



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

약학박사학위논문

질량분석기 기반 대사체학을 이용한
적성병 유래 대사기전의 상관관계
네트워크 분석

Correlation Network Analysis for Deciphering Rust
Disease-Triggered Metabolism Using Mass Spectrometry-
based Metabolomics

2017년 2월

서울대학교 대학원
약학과 약품분석학전공
이 동 규

질량분석기 기반 대사체학을 이용한 적성병
유래 대사기전의 상관관계 네트워크 분석

Correlation Network Analysis for Deciphering Rust Disease-Triggered
Metabolism Using Mass Spectrometry-based Metabolomics

지도교수 권 성 원

이 논문을 약학박사 학위논문으로 제출함

2017년 2월

서울대학교 대학원

약학과 약품분석학전공

이 동 규

이동규의 박사학위논문을 인준함

2017년 2월

위 원 장 박 정 일 (인)

부 위 원 장 양 현 옥 (인)

위 원 권 성 원 (인)

위 원 김 유 선 (인)

위 원 홍 순 선 (인)



국문초록

현재까지의 감염성 질병 연구는, 병원성 균의 침입 방법 및 균이 생성하는 숙주에 대한 자극 물질의 생성에 대해 주로 집중되어 있다. 그러나, 균의 침입에 의한 숙주의 과민반응으로부터 질병의 증상과 같은 병원성 표현형 변화가 유래할 수 있다. 이를 증명하기 위하여, 질량분석기 기반 대사체학을 적용하였으며, 상관관계 네트워크 분석을 이용하여 대사기전 간의 긴밀한 관계를 추적하였다. 대사기전-병원성 연관성 연구를 위해 숙주인 3종의 장미과 식물과 이들에 균체를 형성하여 이들을 메마르게 하는 *Gymnosporangium asiaticum*에 의한 대사기전 변화를 관찰하였다. 질량분석기를 사용하여 다수의 Primary metabolites를 검출하였으며, 다변량통계분석을 통해 얻은 지표물질들을 기반으로 하여 Pathway enrichment를 시행하였다. 이로부터 얻은 상관성 높은 대사기전 사이의 관계를 추적하여, ‘Aminosugar and nucleotide sugar metabolism’의 결핍에 관여되어 있는 대사기전을 발견하였다. 이 대사기전 안에서 주로 당알코올의 축적이 ANM의 Supply metabolites를 결핍시킨다는 것을 PPMCC에 기반한 상관관계 네트워크 분석으로부터 검증하였으며, 교호작용에 대한 관찰로써 Supply metabolites의 결핍과 당알코올의 축적이 연관이 되어 있음을 결론지었다. 이러한 숙주 중심의 대사 반응을 상관관계 네트워크 분석으로 관찰하는 방법으로써 대사기전의 변화 및 그로 인한 결핍이 특정 질병원이 될 수 있음을 보였으며, 대사체학 연구에 적절한 도구로 사용가능함을 확인하였다.

주요어 : 대사체학, 질량분석기, 다변량통계분석, 상관관계 네트워크 분석, 병원성 대사 변화, 당알코올 축적, 대사기질 결핍,

학 번 : 2011-21751

목차

국문초록	i
목차	ii
List of Tables	iii
List of Figures	iv
List of Abbreviations	v
I. 서론	1
II. 재료 및 방법	6
III. 결과 및 고찰	9
IV. 결론	26
V. 참고문헌	27
VI. 부록	31
1. GC-MS spectrum of 34 metabolites	31
2. GC-MS 기반 대사체학	39
3. MZmine	57
4. Multivariate statistical analysis	62
Abstract	73

List of Tables

- Table 1.** List of identified metabolites and the differences between control and parasitized leaves
- Table 2.** Multivariate (VIP value > 1.00) and univariate (p-value < 0.05) statistical parameters of 26 marker metabolites with whole data (n=60) implying importance for discriminating control and parasitized groups.
- Table 3.** Metabolic pathway enrichment results
- Table 4.** Enzymatic pathways about metabolic flows from AMN supplies (fructose, arabinose and galactose) to five sugar alcohols. Enzymes were designated as enzyme entry in KEGG pathway and modules. Enzymes and modules in PPP were colored as green
- Table 5.** Direction, adjusted R square and p-values of the interaction effect between sugars (fructose, arabinose and galactose; dependent variables) and sugar alcohols (ribitol, arabitol, erythritol, mannitol and xylitol; independent variables). Adjusted R square indicates a degree of interaction between them, and the p-value indicates the reliability of the effect.

List of Figures

Figure 1. Taxonomy of *Pucciniales* order

Figure 2. Taxonomy of *Gymnosporangium* genus

Figure 3. Rust disease symptom: aecium formation on pear leaf

Figure 4. Leaf samples of *Crataegus pinnatifida* (control, a; parasitized, b), *Chaenomeles sinensis* (control, c; parasitized, d) and *Pyrus pyrifolia* (control, e; parasitized, f).

Figure 5. Aeciospores of *Gymnosporangium asiaticum* on *Pyrus pyrifolia* (x400)

Figure 6. Control and parasitized leaves of Rosaceae differ in metabolite levels. (a) PLS-DA score plots of CP (diamond, $n = 20$), CS (box, $n = 20$) and PS (triangle, $n = 20$) under the two conditions (parasitism, red; control, green) with PC1 and PC2. Validation parameters are as follows: $R^2X = 0.565$, $R^2Y = 0.963$ and $Q^2 = 0.942$ (b) Scatter column plots of sugar alcohols (C3-C6) and sugars (C5-C6) in control (green) and parasitized (red) leaves divided by carbon number. (c) Scatter column plot with log-scale ratios of C5 sugars and (d) C6 sugars (the centered line represents the mean). (B-D) Sixty samples were assessed, 30 for each group. (e) Logarithmic ratios of metabolites (parasitism/control). Notable metabolites are indicated by asterisks (p -value < 0.05) and a hash sign (VIP value > 1) and ± 2 -fold change by a dashed line ($n = 30$ for each group).

Figure 7. Pathway mapping reveals correlations between ANM supply metabolites and other metabolites. Mapping of metabolic alterations for FMM (a), GM (b) and PGI (c). Metabolites (p -value < 0.05 , marked with an asterisk) between control (green) and parasitized (red) leaves are colored based on their PPMCC values, ranging from -1 (negative correlation, red) to $+1$ (positive correlation, blue), against ANM supply

metabolites (yellow). Values shown are the mean + S.D., with 30 replicates for each group.

Figure 8. Accumulation of sugar alcohols is related to decreases in ANM upstream metabolites. (a) An integrated pathway centered on PPP involving five increased and three decreased metabolites. (b) Heat–maps of correlations between metabolic alterations of the two groups based on intermediates of the integrated pathway (the enzymes in each reaction are indicated in Supplementary Table S4). The color depth denotes the degree of correlation; red color represents the opposite alterations and blue color the equivalent alterations under parasitism.

Figure 9. Abnormal accumulation of C4 and C5 sugar alcohols leads to metabolic deficiency in the ANM pathway. Red arrows indicate increases in metabolites or pathway activity; blue arrows indicate the opposite effect.

List of Abbreviations

FMM: Fructose and mannose metabolism

GM: Galactose metabolism

PGI: Pentose and glucuronate interconversion

PPMCC: Pearson' s product–moment correlation coefficient

PPP: Pentose phosphate pathway

ANM: Aminosugar and nucleotide sugar metabolism

I. 서론

많은 다양한 균의 생존 본능에 따라 자연적으로 기생이란 것이 발생하였으며, 이는 일방향의 필요에 의하여 유지된다. 감염이라는 것은 이러한 기생에 의하여 숙주에서 영양분 또는 몇몇 기능을 취할 때 숙주의 정상적인 생활에 문제가 나타나는 현상이며 이는 인간과 Microbiome, 식물과 미생물들간의 관계 등 대다수의 종에서 나타나는 현상이다. 이 감염이라는 현상으로 인하여 많은 생명체들이 생존을 위협받고 있으며, 인간의 경우도 다양한 미생물로 인한 질병을 갖고 있음이 알려져 있다. 따라서 감염에 대한 연구는 생명 연구의 주요 분야이며, 다양한 질병의 치료를 위하여 반드시 필요한 항목이다.

감염 현상을 연구하기 위한 생물학적 연구는 다양하게 이루어져 있다 [1]. 감염균이 숙주에 기생하기 시작하게 되는 Factor와 이들 사이의 신호전달물질 등에 대해서는 물론이고, 이러한 신호전달물질이 숙주에서 어떤 Receptor에서 수용을 하여 숙주를 변화시키는 지도 연구되어 있다 [2-4]. 그러나 이들은 질병의 원인을 기생하는 균을 중심으로 연구하였으며, 기생균의 생화학적 명령에 따른 숙주의 반응에 대해서는 연구가 미진하다.

따라서 기생 균의 침입에 따른 숙주의 변화를 생화학적 변화의 말단에 위치하는 대사체 수준에서 관찰하기 위하여 질량분석기에 기반한 대사체학을 적용하였다. 기생 균이 자신들이 유리한 방향으로 숙주의 대사를 변화시킨다는 것은 알려져 있다 [5-7]. 따라서 병원균에 의해 유도된 대사의 변화가 질병의 현상과 관련이 있을 가능성은 농후하다. 그렇기에 다양한 감염 후 변화의 Phenotype과 대사기전과의 관계를 연구하게 될 경우 왜 기생 균이 특정 숙주의 정상적인 생존을 방해하는지에 대한 원인을 찾을 수

있을 것이다.

그러나 현재 질병의 원인을 밝히기 위한 대사체학의 연구는 대체로 병원성 유전체 및 단백질을 발굴하기 위한 기반 연구로 여기는 경우가 대부분이다. 병원성과 관련된 지표대사체와 연관성 있는 유전체 또는 단백질체의 발현을 확인하여 그 기능의 이상을 병원성으로 여긴다. 그러나 숙주의 대사기전은 수만 가지 대사체의 네트워크로 구성되어 있기에 한 두개의 유전체 또는 단백질체의 발현에서 그 질병 요인을 찾는 것은 어렵다. 따라서 이 연구에서 실타래처럼 엉켜있는 대사기전 네트워크간의 상관관계로부터 특정 감염에 따른 질병 발생의 원인, 즉 감염에 의한 다양한 변형의 원인을 밝혀내는 방법을 적용해보았다.

그 적용을 위하여 해당 논문에서는 적성병을 사용하였다. 적성병은 Fungi의 한 종류이며 Basidiomycota phylum, Pucciniomycetes class 중 the order Pucciniales에 속하는 총 168개의 genera, 약 7000개의 species에 의해 발생하는 식물 병원균이다 (Figure 1). 이들은 각자의 숙주에 spore를 형성하는 녹을 형성하며, reproduction을 위하여 spermatogonia, aecia, uredinia, telia, basidia등의 다양한 단계의 reproduction을 거친다.

그 중에서도 해당 논문에서는 *Gymnosporangium* genus 균에 의해 감염되는 적성병을 연구하였다 (Figure 2). 이 균의 경우는 *Juniperus* species와 Rosaceae family를 모두 숙주로하여 이종기생하는 균이다. 이 균은 겨울을 월동하며 *Juniperus*의 petiole 등에서 yellow gall을 형성하며, 봄이 되어 비가 내리면서 수분을 머금은 균이 다시 활동하여 spore를 공기중으로 날려보낸다. 이 때 반경 1~2 km에 있는 장미과 식물에 spore가 형성되면서 aecium을 형성하여 장미과 식물에 해를 가한다 (Figure 3).

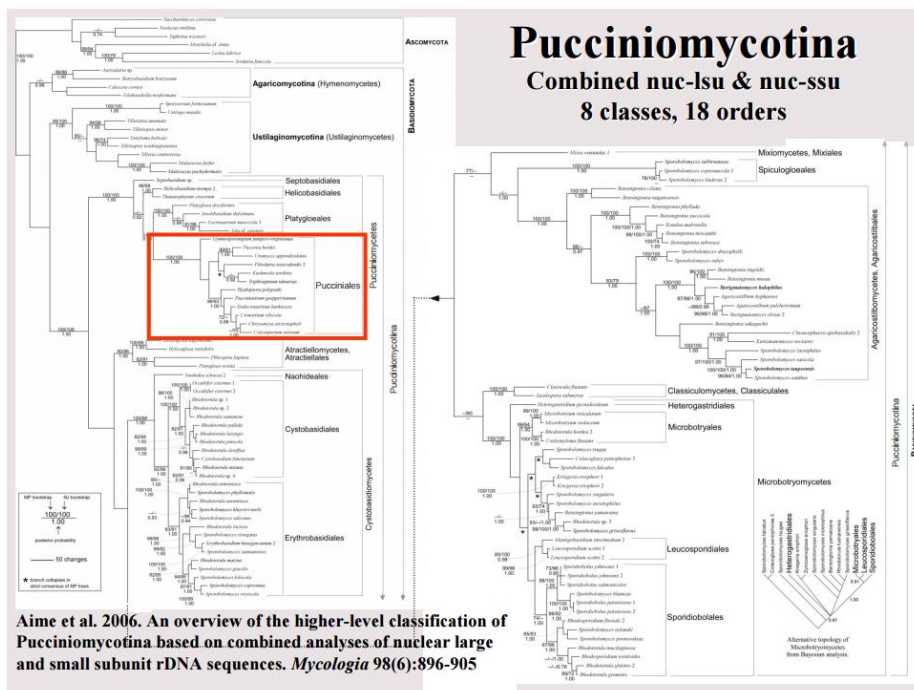


Figure 1. Taxonomy of *Pucciniales* order (reference: <https://www.plantmanagementnetwork.org/infocenter/topic/soybeanrust/2009/presentations/Aime.pdf>)

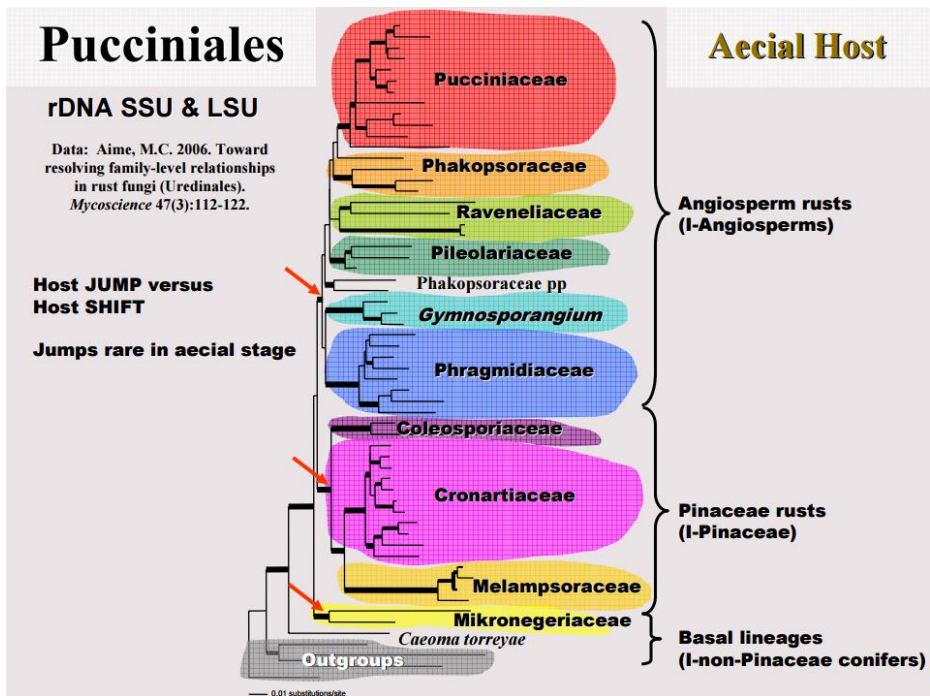


Figure 2. Taxonomy of *Gymnosporangium* genus (reference: <https://www.plantmanagementnetwork.org/infocenter/topic/soybeanrust/2009/presentations/Aime.pdf>)



Figure 3. Rust disease symptom: aecium formation on pear leaf (reference: <http://blog.naver.com/yjs3376/220105052213>)

그런데 이러한 균의 감염 경로 등에 대해서는 연구가 되어 있으나, 왜 해당 균이 장미과 식물에 접촉했을 때 aecium이 생성되고 숙주를 메마르게 하는지에 대해서는 연구가 되어있지 않다. 따라서 적성병의 병리 기전을 숙주의 차원에서 밝히기 위하여, 위에서 언급한 대사기전 네트워크 분석을 적용해보았다. 이로부터 식물의 병, 더 나아가 동물 및 인간의 질병의 요인을 숙주차원에서 밝힐 수 있는 대사체학적 접근법을 보여주고자 한다. 또한 단백질, 유전체 단계에서가 아닌, 대사기전 네트워크 차원에서의 불균형 등이 질병의 요인이 될 수도 있음을 해당 논문에서 밝히고자 한다.

II. 재료 및 방법

1. 시료 준비

각각 다른 장소에 있는 *Crataegus pinnatifida* (산사나무), *Chaenomeles sinensis* (모과나무), *Pyrus pyrifolia* (배나무)를 Control과 Parasitized (aecial stage) 그룹별로 각각 10개씩, 총 60개의 시료를 수집하였다. 모든 나무의 잎, 열매, 잎꼭지에 국소적으로 부풀어있고 그 위에 녹이 있는 것으로 보아, 이는 적성병의 단계 중 aecial stage에 있음을 확인하였다. 대사체의 변화의 패턴이 달라질 수 있음을 고려하여, 1년 뒤 동일한 장소에서 각 그룹별로 3개의 잎을 추가로 채집하였다. 모든 샘플들은 수집 후 즉각적으로 -40°C 에서 보관되었고, -70°C deep freezer에서 보관되었다. 잎들은 녹, 부어오른 부위, 그 외의 부위 3가지로 나뉘어 구분되었으며, 잎꼭지, 중맥, 잎맥은 모두 제거되었다. 대사체 추출 시 잎에 남아있는 균은 모두 증류수를 이용하여 제거하였으며, 24시간의 동결건조 후, 체를 통과한 시료를 이용해 대사체를 추출하였다.

2. 균 동정

Crataegus pinnatifida, *Chaenomeles sinensis*, *Pyrus pyrifolia*에서 총 9개의 *Gymnosporangium* specimens을 수집하였다. Compound light microscope가 균의 모양, peridium, peridial cell 크기와 같은 특징을 관찰하는데 사용되었다. 400배로 확대하여 aeciospore의 모양, 크기, 세포벽 두께를 확인하였으며, 각각의 specimen 별로 30개의 spore를 관찰하였다. 외형적 특징을

기준에 게재된 *Gymnosporangium*의 묘사와 비교해보았을 때, 이 적성병 균은 *Gymnosporangium asiaticum*으로 동정 되었다 [8].

3. 대사체 프로파일링

파우더 상태의 잎 20 mg이 1:2.5:1 chloroform:methanol:water 용액에서 30분간 Sonication 하였다. C17:0 fatty acid는 internal standard 로써 각 샘플에 동일한 농도로 첨가하였다. 13,000 g에서 5분간 centrifuge 후 상층액은 N₂ purge 를 통해 용매를 휘발 시켰다. 추출된 대사체들은 methoxyamination 과 trimethylsilylation을 거쳤다 [9, 10]. 이들은 DB-5 컬럼 (30 m X 0.25 mm, 0.25 μ m thickness) 이 연결된 GC-MS에서 분리 및 검출되었다. 유도체화 된 대사체들은 40.33 분의 temperature program을 거쳐 electron impact 이온화 방식을 통해 MS detector로 유입되었다. 검출 범위는 40~600 m/z 였으며, retention index solution (C7-C40 alkane mixture)로부터 RI를 도출하여 대사체의 identification에 사용되었다. 각각의 대사체 스펙트럼은 AMDIS program으로 deconvolution 되었으며, 스펙트럼이 유사한 당류의 경우는 스탠다드 데이터를 획득하여 추가적으로 확인하였다.

4. 데이터 프로세싱

60 샘플의 데이터는 MZmine을 이용하여 alignment 하였다 [11, 12]. 정렬된 bucket table은 SIMCA-P⁺를 이용하여 주성분분석 (principal component analysis), 최소자승판별법 (partial least square-discriminant analysis) 결과를 획득하는데 사용되었다.

단변량 및 다변량 지표 물질 판별값으로써 각각 p-value 와 variable importance for projection (VIP) value를 획득하였으며, 대사체간의 상호관계 (correlation)을 관찰하기 위해서 Pearson's product-moment correlation coefficient가 측정되었다 [13]. 각 대사체들의 pathway mapping에는 VANTED software가 사용되었으며 [14], 대사기전은 KEGG pathway에서 KGML file 형식으로 획득하였다 [15]. Pathway enrichment는 MBrole을 이용하였으며, 교호작용 (interaction effect) 계산에는 SPSS 가 사용되었다.

III. 결과 및 고찰

1. 균 동정

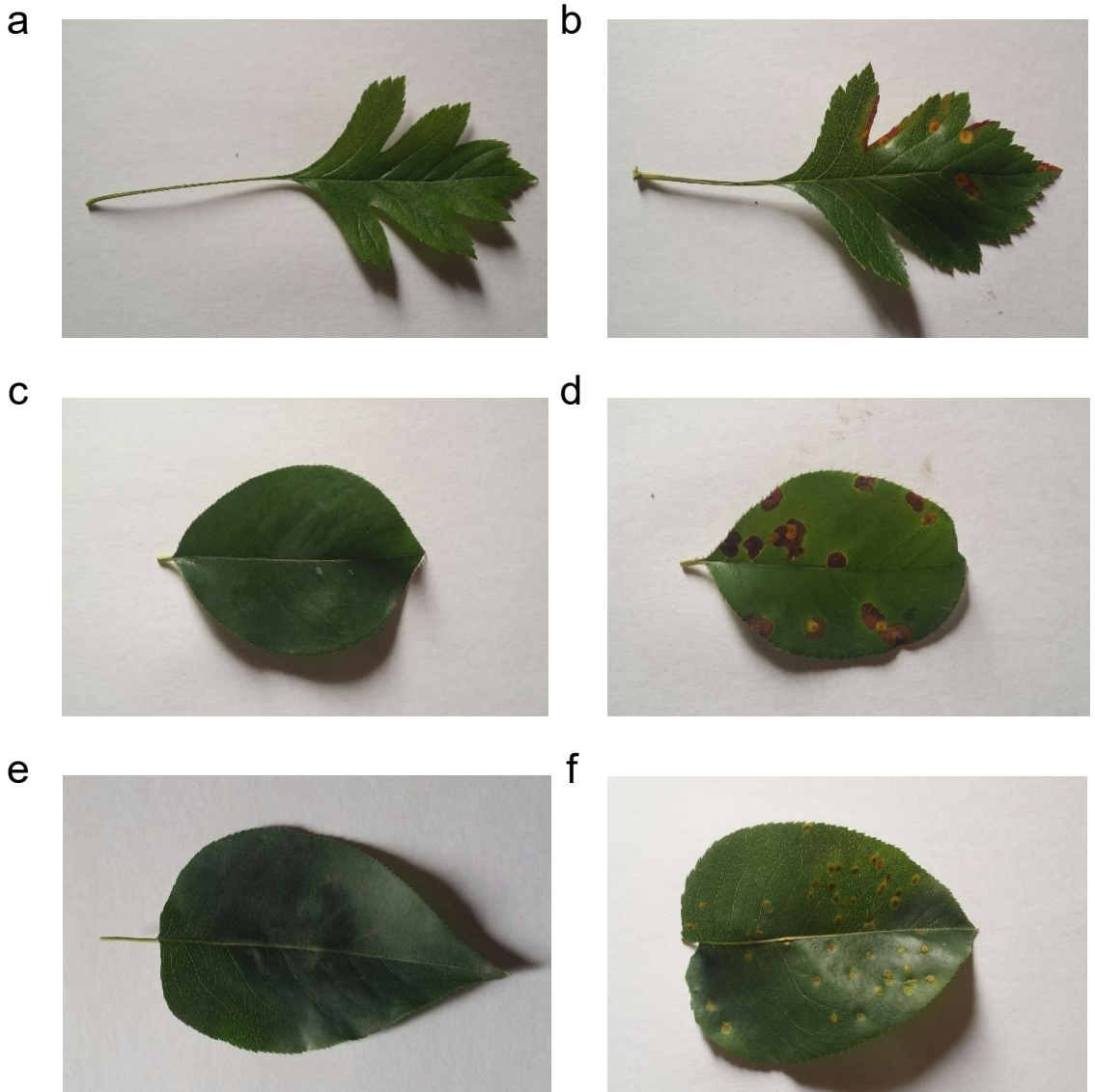


Figure 4. Leaf samples of *Crataegus pinnatifida* (control, a; parasitized, b), *Chaenomeles sinensis* (control, c; parasitized, d) and *Pyrus pyrifolia* (control, e; parasitized, f).

우리는 정상 군과 감염 군의 시료를 Rosaceae 에 포함되는 3가지 종인 *Crataegus pinnatifida* (산사나무, 이하 CP), *Chaenomeles sinensis* (모과나무, 이하 CS) and *Pyrus pyrifolia* (배나무, 이하 PS) 에서 획득하였다 (Figure 4). 이들은 deformity와 fungal aecia를 갖고 있었으며, Morphological characteristic을 보았을 때 적성병에 해당하였다. 구체적으로 어떤 종의 군에 감염되었는지를 확인하기 위해서 aeciospore를 관찰하였으며, 형태학적 특징으로부터 *Gymnosporangium asiaticum*에 해당하는 것을 확인하였다 (Figure 5) [16]. 이를 생화학적인 방법으로 규명하고자 DNA sequence를 jPHYDIT 을 이용한 pairwise alignment한 후에, GenBank database에 있는 해당 군의 sequence (FJ848744, FJ848745, FJ848746, 630 bp matching)와 비교하였다 [16-19]. 이는 대한민국에서 발견되었다고 논문에 게재된 specimens와 일치함을 확인하였다.

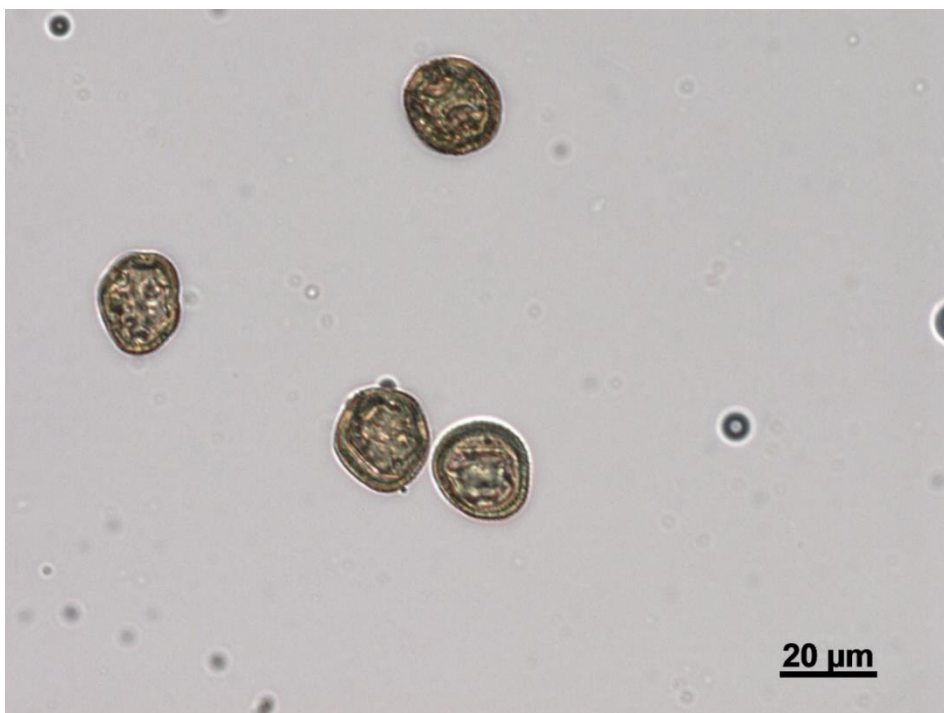


Figure 5. Aeciospores of *Gymnosporangium asiaticum* on *Pyrus pyrifolia* (x400)

2. 감염에 따른 대사체의 변화 관찰

균에 의해 감염되었을 때 나타나는 대사체의 패턴을 찾기 위해서 각 3가지 종의 식물에서 대사체를 획득하였다 (Table 1). 대사체의 identification을 위하여 유도체화 된 대사체들의 retention time과 fragment ion 구성 비율을 확인하였으며, 이를 NIST database (NIST08 version)과 비교하였다. 또한 유사 구조를 지닌 당류의 identification을 확실히 하기 위하여 standard를 이용해 유도체화 된 대사체의 retention index와 비교 확인하였다. 결과적으로, 총 34개의 대사체를 확인하였으며, organic acid 7종류, amino acid 2종류, fatty acid 2종류, sugar acid 3종류, 당알코올 8종류, 당 12종류를 관찰하였다. 이 모든 대사체 유도체화물의 Mass spectrum은 IV. 부록에 추가하였다.

Table 1. List of identified metabolites and the differences between control and parasitized leaves

Retention time (min)	Compound name	Fragment ion (derivatized compound) ^a	p-value	Fold change	Identification NIST	STD
8.04	Phosphoric acid, 3TMS	298(100), 73(60), 313(18), 45(11)	0.13	0.73	○	
8.08	Glycerol, 3TMS	73(100), 147(78), 205(43), 117(32), 103(31)	0.18	1.47	○	○
8.73	Succinic acid, 2TMS	147(100), 73(52), 75(19), 148(16), 246(10)	3.0×10^{-4}	0.54	○	
9.25	Glyceric acid, 3TMS	73(100), 147(61), 189(31), 292(23), 102(17)	6.5×10^{-13}	0.53	○	
9.41	Fumaric acid, 2TMS	245(100), 73(57), 147(41), 143(23)	NM		○	
9.84	Serine, 3TMS	73(100), 204(95), 217(53), 100(27), 147(20)	NM		○	
10.43	Threonine, 3TMS	73(100), 117(48), 218(41), 57(22), 147(17)	NM		○	
12.16	Citramalic acid, 3TMS	73(100), 147(53), 247(41), 75(23), 115(15)	2.3×10^{-4}	0.33	○	
12.47	Malic acid, 2TMS	73(100), 147(53), 233(18)	0.05	0.87	○	
13.04	Erythritol, 4TMS	73(100), 147(57), 216(51), 103(34), 205(28)	1.3×10^{-13}	2.85	○	○
13.10	Aspartic acid, 3TMS	232(100), 73(98), 100(32), 147(22), 218(16)	NM		○	
13.73	Erythronic acid, 4TMS	73(100), 147(51), 292(24), 117(14), 103(11)	NM		○	
14.69	3-H-3-MP acid, 3TMS ^b	73(100), 147(47), 247(31), 115(18), 231(18)	NM		○	
15.27	Xylonic acid, 3TMS	73(100), 217(29), 147(27), 129(18), 102(11)	0.93	0.99	○	
16.00	Xylose, 4TMS, 1 MO	73(100), 103(77), 217(49), 307(32), 147(28)	0.018	1.50	○	
16.12	Arabinose, 4TMS	73(100), 103(88), 217(43), 307(29), 147(25)	3.0×10^{-8}	0.50	○	○
16.40	Ribose, 4TMS, 1 MO	73(100), 103(80), 217(44), 307(28), 147(24)	0.25	0.88	○	○
17.00	Xylitol, 5TMS	73(100), 217(64), 103(42), 147(40), 205(21)	0.015	1.49	○	○
17.21	Arabitol, 5TMS	73(100), 217(63), 103(44), 147(41), 205(22)	3.9×10^{-23}	11.39	○	○
17.29	Ribitol, 5TMS	73(100), 217(57), 147(43), 103(40), 205(27)	2.5×10^{-7}	2.43	○	○
17.92	Arabinofuranose, 4TMS	217(100), 73(92), 147(31), 232(10)	NM		○	○
18.01	Ribonic acid, 5TMS	73(100), 292(45), 147(36), 103(30), 217(29)	0.64	0.93	○	
18.23	Idonic acid, 4TMS	73(100), 147(36), 217(28), 103(24), 436(20)	8.2×10^{-4}	0.51	○	
18.82	Citric acid, 4TMS	73(100), 273(71), 147(59), 347(17), 375(14)	0.017	1.18	○	
19.91	Sorbose, 5TMS	73(100), 103(76), 217(52), 307(32), 147(24)	0.17	0.92	○	○
20.09	Mannose, 5TMS, 1 MO	73(100), 319(48), 205(41), 147(39), 160(26)	0.24	0.92	○	○
20.37	Fructose, 5TMS, 1 MO	73(100), 103(73), 217(53), 307(33), 147(24)	9.0×10^{-8}	0.46	○	○
20.65	Galactose, 5TMS, 1 MO	73(100), 319(49), 205(40), 147(39), 217(21)	1.0×10^{-10}	0.39	○	○
20.93	Mannitol, 6TMS	73(100), 319(56), 147(43), 205(39), 217(35)	1.7×10^{-17}	2.86	○	○
21.07	Sorbitol, 6TMS	73(100), 319(53), 147(42), 205(36), 103(29)	3.6×10^{-13}	0.35	○	○
22.10	Gluconic acid, 6TMS	73(100), 147(36), 333(28), 292(25), 103(15)	0.79	0.94	○	
23.39	myo-Inositol, 6TMS	73(100), 217(60), 305(58), 147(44), 318(31)	0.044	1.30	○	○
23.79	Glucose, 5TMS, 1 MO	73(100), 319(63), 205(38), 147(36), 103(21)	0.84	0.97	○	○
24.61	Linoleic acid, 1TMS	75(100), 73(97), 67(77), 81(70), 55(54)	NM		○	
24.70	Oleic acid, 1TMS	73(100), 117(94), 75(90), 129(64), 55(56)	1.54×10^{-6}	0.61	○	
25.06	Stearic acid, 1TMS	117(100), 73(75), 75(56), 341(55), 132(54)	0.061	0.81	○	
31.10	Sucrose, 8TMS	73(100), 361(88), 216(37), 147(25), 103(22)	0.32	0.82	○	○
32.84	Maltose, 8TMS	73(100), 361(86), 204(59), 147(34), 217(31)	0.007	1.70	○	○
33.97	Melibiose, 8TMS	204(100), 73(66), 217(29), 205(28), 361(27)	0.041	0.76	○	○

얻어진 대사체 프로파일 데이터들을 Alignment 하여 Bucket table 을 얻었으며, 감염군에 의해 유발된 대사체 구성의 변화를 찾기 위하여 이들을 이용하여 principal component analysis (PCA) score plot 을 획득하였다. CP (PC1, explaining 47.4%), CS (PC1, explaining 45.9%), PS (PC1, explaining 54.5%) 의 scatter plot 을 얻었으며, 이들은 R^2_{cum} (순서대로 0.751, 0.787, 0.830)과 Q^2_{cum} (순서대로 0.475, 0.751, 0.686) 의 validation parameter 를 가지며 정상 군과 감염 군의 판별이 적합했음을 확인하였다. 그러나 총 60 개의 모든 데이터를 이용하여 PCA score plot 을 얻은 결과, 정상군과 감염 군의 판별보다는 각 중별로의 대사체 프로파일 차이를 더욱 부각시켰다는 것을 확인하였다. 이는 unsupervised method 인 PCA 의 특성상 가장 두드러지는 판별 기준을 설정하였기 때문이며, 이로부터는 정상 군과 감염 군을 판별할 수 있는 지표대사체를 찾는 것이 힘들다는 것을 확인하였다. 따라서 supervised method 의 하나인 partial least square-discriminant analysis (PLS-DA)를 사용하였으며, 이로부터 각 대사체가 정상 군, 감염 군을 판별하는데 중요한 정도 확인해주는 variable importance for projection (VIP) value 를 얻었다 (Figure 6a). 또한 다변량통계인 PLS-DA 외에, 단변량 통계분석으로써 유의성을 파악하는 t-test 를 통해 얻은 p-value 를 얻었다 (Table 2).

Table 2. Multivariate (VIP value > 1.00) and univariate (p-value < 0.05) statistical parameters of 26 marker metabolites with whole data (n=60) implying importance for discriminating control and parasitized groups.

Metabolites	VIP value	p-value
Erythritol	2.35	1.3×10^{-13}
Arabitol	2.34	3.9×10^{-23}
Galactose	1.89	1.0×10^{-10}
Citric acid	1.65	0.017
Maltose	1.50	0.007
myo-Inositol	1.47	0.044
Mannitol	1.45	1.7×10^{-17}
Arabinose	1.40	3.0×10^{-8}
Xylitol	1.40	0.015
Malic acid	1.39	0.05
Glucose	1.37	0.84
Ribitol	1.33	2.5×10^{-7}
Glycerol	1.28	0.18
Gluconic acid	1.26	0.79
Phosphoric acid	1.25	0.13
Fructose	1.22	9.0×10^{-8}
Mannose	1.08	0.24
Ribose	1.03	0.25
Sorbitol	1.01	3.6×10^{-13}
Sorbose	1.00	0.17
Succinic acid	<1.00	3.0×10^{-4}
Glyceric acid	<1.00	6.5×10^{-13}
Citramalic acid	<1.00	2.3×10^{-4}
Oleic acid	<1.00	1.54×10^{-6}
Melibiose	<1.00	0.041
Xylose	<1.00	0.018

위에서 얻은 지표대사체 관련 수치들과 각 대사체의 fold change로부터 감염되었을 때 변하는 지표 대사체를 관찰하였다 (Figure 6e).

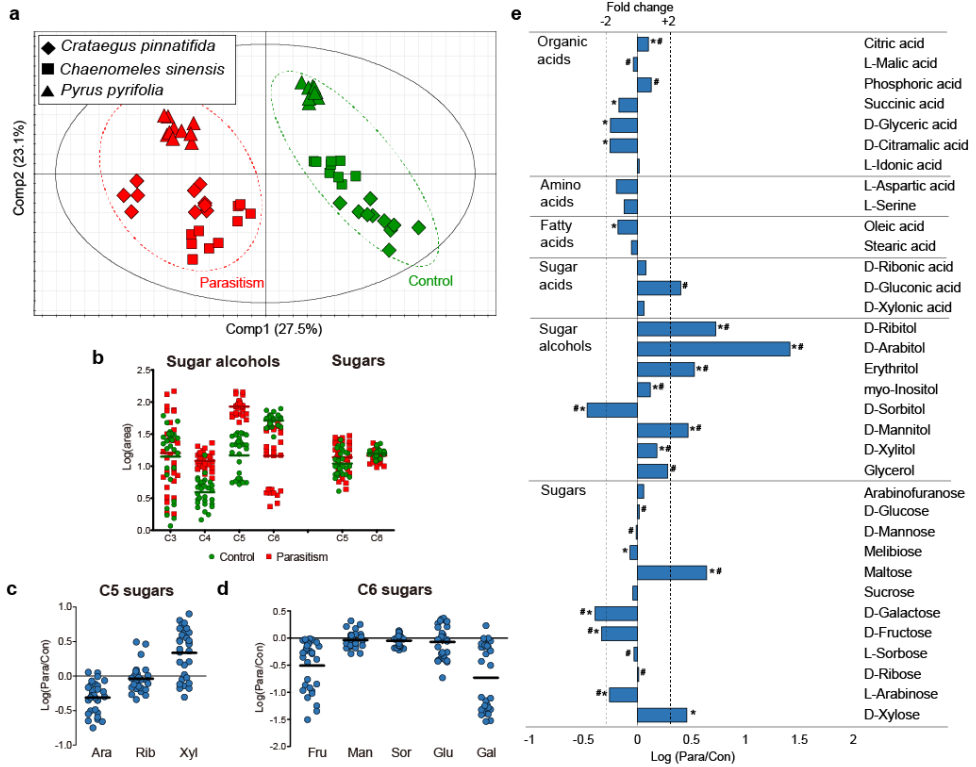


Figure 6. Control and parasitized leaves of Rosaceae differ in metabolite levels. (a) PLS-DA score plots of CP (diamond, n = 20), CS (box, n = 20) and PS (triangle, n = 20) under the two conditions (parasitism, red; control, green) with PC1 and PC2. Validation parameters are as follows: $R^2X = 0.565$, $R^2Y = 0.963$ and $Q^2 = 0.942$ (b) Scatter column plots of sugar alcohols (C3-C6) and sugars (C5-C6) in control (green) and parasitized (red) leaves divided by carbon number. (c) Scatter column plot with log-scale ratios of C5 sugars and (d) C6 sugars (the centered line represents the mean). (B-D) Sixty samples were assessed, 30 for each group. (e) Logarithmic ratios of metabolites (parasitism/control). Notable metabolites are indicated by asterisks (p-value < 0.05) and a hash sign (VIP value > 1) and ± 2 -fold change by a dashed line (n = 30 for each group).

VIP value 는 1.00 이상을 기준으로, p-value 는 0.05 이하를 기준으로 하여 중요한 대사체를 선별하였으며, 이로부터 총 26 개의

대사체들이 감염되었을 때 변화한다는 것을 확인하였다. 거의 모든 대사체가 변하는 것이므로, 숙주내 대사를 통틀어서 변화시킨다는 것을 알 수 있었다. 다양한 지표 물질 중 당 및 당알코올류에 가장 많은 지표대사체가 있었으며, fold change 는 sugar alcohol 이 당에 비해서 더 큰 양상을 보이고 있었다. 따라서 감염되었을 때 숙주 대사의 변화는 당과 당알코올이 주된 비율을 차지하고 있으며, 다양한 증상에 이들이 관여할 수 있다는 것을 보여주었다.

3. 균에 의해 감염되었을 때 당과 당알코올의 규제

감염되었을 때 가장 크게 영향을 받은 대사체인 당과 당알코올을 탄소수에 따라서 구분 하였으며, 이를 column-scattered 형식으로 표현하였다 (Figure 4b). 당알코올을 보면, C3 의 경우는 유의성 있는 차이를 나타내지 않았으나, C4 (erythritol), C5 (ribitol, arabitol, xylitol) 의 경우는 감염 균에서 높게 나타나는 것을 보였으며, 이는 위의 다변량 및 단변량 통계분석의 지표대사체 판별 결과와 상응한다. 오히려 C6 (mannitol, sorbitol, inositol)의 경우는 반대로 감염 균에서 감소하는 것을 보였다. 당의 경우는 모든 당을 모아서 보았을 경우 차이를 나타내지 않았으나, Figure 4c 와 같이 각 당으로 나누어서 보았을 경우 arabinose, fructose, galactose 가 감소하는 것이 확인되었다. 결론적으로 균의 감염은 숙주의 C4, C5 당알코올을 축적하며, C5, C6 당을 감소시키는 쪽으로 대사를 변화시키는 것을 확인하였다.

4. 병원성 대사기전 상관관계 분석

지표대사체로 선정된 26 개의 대사체를 이용하여 이들과 관계가 있는 대사기전을 밝혀내었다. MBrole 을 이용하여 Pathway enrichment 를 해보았으며, 대사기전에 대한 정보는 KEGG pathway 에 수록된 정보를 사용하였다 (Table 3).

Table 3. Metabolic pathway enrichment results.

Metabolic pathway	Adjusted p-value	Total metabolites included	Matched metabolites	Matched/Total (%)
Pentose and glucuronate interconversions	2.08×10^{-04}	53	5	27.8
Galactose metabolism	9.06×10^{-04}	41	4	22.2
Fructose and mannose metabolism	0.015	48	3	16.7
Starch and sucrose metabolism	0.015	50	3	16.7
Glyoxylate and dicarboxylate metabolism	0.015	44	3	16.7
Biosynthesis of plant hormones	0.026	68	3	16.7
Citrate cycle (TCA cycle)	0.026	20	2	11.1
Biosynthesis of alkaloids derived from histidine and purine	0.056	35	2	11.1
Biosynthesis of phenylpropanoids	0.056	97	3	16.7
Biosynthesis of alkaloids derived from terpenoid and polyketide	0.084	48	2	11.1
Ascorbate and aldarate metabolism	0.084	47	2	11.1
Biosynthesis of alkaloids derived from shikimate pathway	0.102	138	3	16.7
Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid	0.122	67	2	11.1
Amino sugar and nucleotide sugar metabolism	0.181	87	2	11.1
Biosynthesis of terpenoids and steroids	0.208	98	2	11.1

Adjusted p-value 가 0.05 이하이며, matched/total percentage 가 높은 대사기전이 감염군에서의 변화와 관련도가 높은 대사기전이며, 7 개의 대사기전이 의미가 높았다. 따라서 그들 7 개 대사기전간의 연관성을 찾아보았다. 이 때, starch and sucrose metabolism 과 glyoxylate and dicarboxylate metabolism 을 제외하고 모든 대사기전이 aminosugar and nucleotide sugar metabolism (ANM) pathway 와 관련있음을 확인하였다. Pentose and glucuronate interconversion (PGI), galactose metabolism (GM), fructose and mannose metabolism

(FMM) 들의 product 가 ANM 대사기전의 substrate 공급으로써 사용되고 있다는 것을 확인하였다. 놀랍게도 ANM pathway 의 upstream metabolites 는 외부에서 공급되어야 하는 N-acetylmuramate, N-acetyl-D-glucosamine 을 제외하고는 모든 대사체가 위의 3 가지 대사기전으로부터 공급된다는 점이다.

따라서 감염 시 균에 의해 유도되는 숙주 대사기전의 변화 및 이들간의 상호관계를 확인하는 네트워크 분석을 진행하였다. 일단 ANM 의 substrate 이자 위 3 가지 대사기전의 product 인 대사체들을 중심으로 pathway mapping 을 진행하였다 (Figure 7) [20].

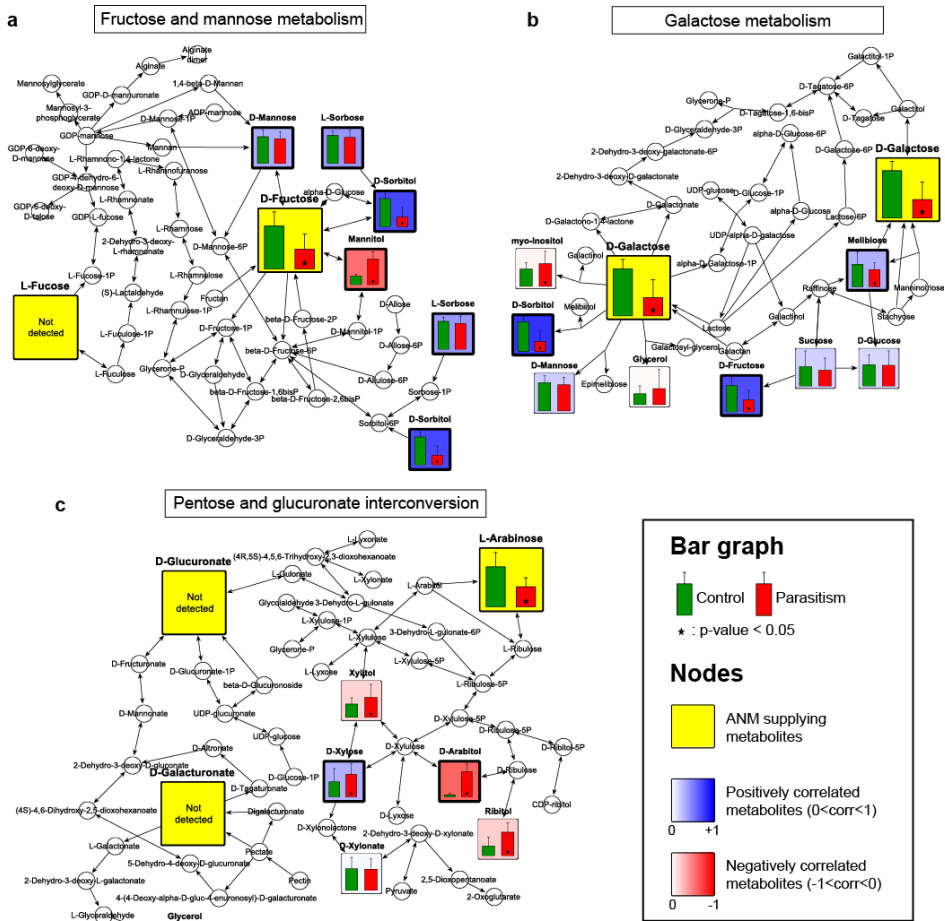


Figure 7. Pathway mapping reveals correlations between ANM supply metabolites and other metabolites. Mapping of metabolic alterations for FMM (a), GM (b) and PGI (c). Metabolites (p-value < 0.05, marked with an asterisk) between control (green) and parasitized (red) leaves are colored based on their PPMCC values, ranging from -1 (negative correlation, red) to +1 (positive correlation, blue), against ANM supply metabolites (yellow). Values shown are the mean + S.D., with 30 replicates for each group.

각각의 대사기전에서 노란색으로 표시된 대사체가 ANM supply metabolites 이며, 각각의 기전에 not detected 로 표시된 fucose, glucuronate, galacturonate 는 스탠다드를 통한 유도체화로서 검출을 확인하였으나, 실제 시료에는 함유되지 않은 것으로 확인하였다. 이들은 모두 유의성 있게 감소하는 것을 보여주었으며,

이들의 감소와 상관관계를 나타내는 대사체의 증감을 찾고자 PPMCC value 를 빨간색 (음의 상관관계, $0 \sim -1$)과 파란색 (양의 상관관계, $0 \sim 1$)로써 pathway mapping 에 적용하였다. 이때 음의 상관관계를 갖는 경우가 ANM supply metabolites 의 감소를 유도하는 대사체의 축적이다. 이로부터 galactose 의 감소에는 galactose metabolism 내 대사기전에 관계 있는 경우가 없음을 확인하였으며, 다른 두 대사기전의 arabinose 와 fructose 의 감소에는 유의성 있게 축적되는 것으로 확인되었던 당알코올 (mannitol, arabitol, xylitol, ribitol)이 높은 음의 상관관계를 보이면서 크게 관여하고 있다는 것을 밝혀냈다. 이로부터 당알코올과 ANM supply metabolites 에 긴밀한 관계가 있음을 추론할 수 있다.

놀랍게도, 이 축적되는 당알코올, 그리고 감소하는 당들은 모두 pentose phosphate pathway 에서 서로 연결되어 있음을 확인하였다[21-24]. KEGG database 에 근거하여 보았을 때, 이들은 모두 각각의 phosphate form 으로써 대사체 네트워크를 형성하고 있다. 이 긴밀한 연결은 Figure 8a 및 Table 4 에서 확인할 수 있다.

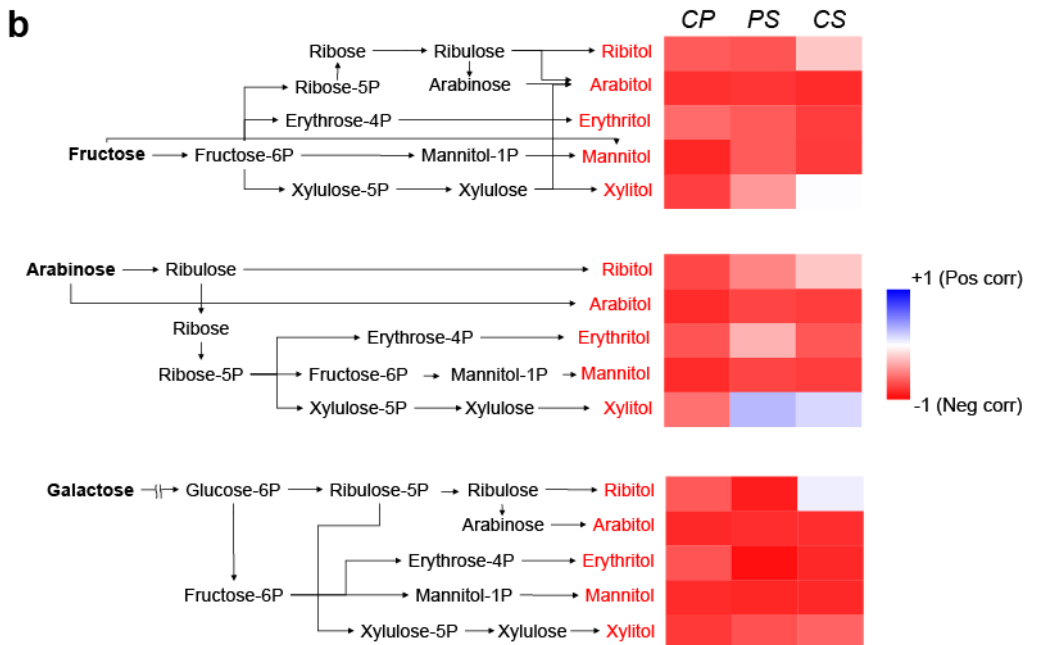
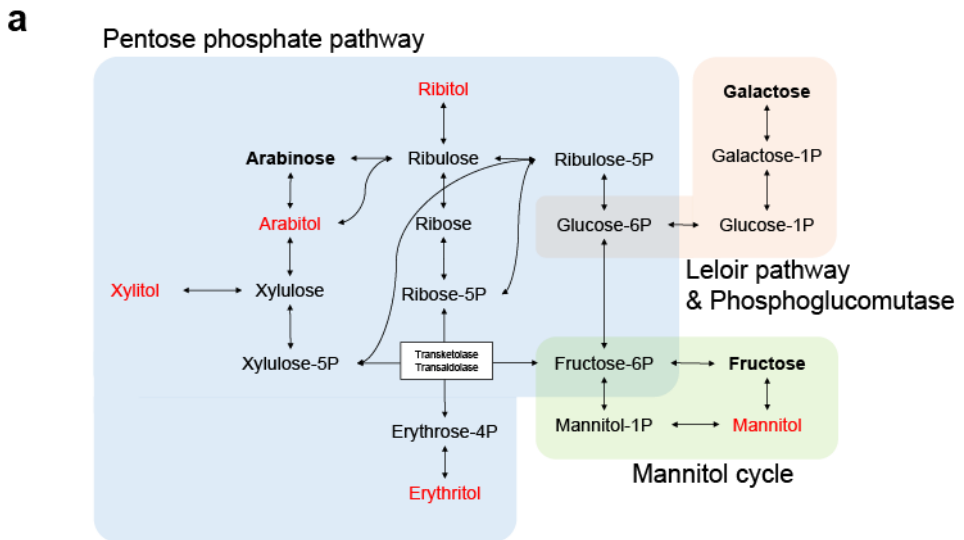


Figure 8. Accumulation of sugar alcohols is related to decreases in ANM upstream metabolites. (a) An integrated pathway centered on PPP involving five increased and three decreased metabolites. (b) Heat-maps of correlations between metabolic alterations of the two groups based on intermediates of the integrated pathway (the enzymes in each reaction are indicated in Supplementary Table S4). The color depth denotes the degree of correlation; red color represents the opposite alterations and blue color the equivalent alterations under parasitism.

Table 4. Enzymatic pathways about metabolic flows from AMN supplies (fructose, arabinose and galactose) to five sugar alcohols. Enzymes were designated as enzyme entry in KEGG pathway and modules. Enzymes and modules in PPP were colored as green

Fructose	Ribitol	Fructose $\xrightarrow{EC:2.7.1.1}$ Fructose-6P $\xrightarrow{M00007}$ Ribose-5P $\xrightarrow{EC:2.7.1.15}$ Ribose $\xrightarrow{EC:5.3.1.20}$ Ribulose $\xrightarrow{EC:1.1.1.56}$ Ribitol
	Arabitol	Fructose $\xrightarrow{EC:2.7.1.1}$ Fructose-6P $\xrightarrow{M00007}$ Ribose-5P $\xrightarrow{EC:2.7.1.15}$ Ribose $\xrightarrow{EC:5.3.1.20}$ Ribulose $\xrightarrow{EC:1.1.1.13}$ Arabitol Fructose $\xrightarrow{EC:2.2.1.1}$ Xylulose-5P $\xrightarrow{EC:2.7.1.17}$ Xylulose $\xrightarrow{EC:1.1.1.287}$ Arabitol Ribulose $\xrightarrow{EC:5.3.1.4}$ Arabinose $\xrightarrow{EC:1.1.1.21}$ Arabitol
	Erythritol	Fructose $\xrightarrow{EC:2.7.1.1}$ Fructose-6P $\xrightarrow{EC:2.2.1.1}$ Erythrose-4P $\xrightarrow{EC:2.7.1.27}$ Erythritol
	Mannitol	Fructose $\xrightarrow{EC:2.7.1.1}$ Fructose-6P $\xrightarrow{EC:1.1.1.67 \text{ or } 1.1.1.138}$ Mannitol-1P $\xrightarrow{EC:3.1.3.22}$ Mannitol
	Xylitol	Fructose $\xrightarrow{EC:2.7.1.1}$ Fructose-6P $\xrightarrow{EC:2.2.1.1}$ Xylulose-5P $\xrightarrow{EC:2.7.1.17}$ Xylulose $\xrightarrow{EC:1.1.1.15, EC:1.1.1.10, EC:1.1.1.9}$ Xylitol
Arabinose	Ribitol	Arabinose $\xrightarrow{EC:5.3.1.4}$ Ribulose $\xrightarrow{EC:1.1.1.56}$ Ribitol
	Arabitol	Arabinose $\xrightarrow{EC:1.1.1.21}$ Arabitol
	Erythritol	Arabinose $\xrightarrow{EC:5.3.1.4}$ Ribulose $\xrightarrow{EC:5.3.1.20}$ Ribose $\xrightarrow{EC:2.7.1.15}$ Ribose-5P $\xrightarrow{EC:2.2.1.2, EC:2.2.1.1}$ Erythrose-4P $\xrightarrow{EC:2.7.1.27}$ Erythritol
	Mannitol	Arabinose $\xrightarrow{EC:5.3.1.4}$ Ribulose $\xrightarrow{EC:5.3.1.20}$ Ribose $\xrightarrow{EC:2.7.1.15}$ Ribose-5P $\xrightarrow{M00007}$ Fructose-6P $\xrightarrow{EC:1.1.1.17}$ Mannitol-1P $\xrightarrow{EC:3.1.3.22}$ Mannitol
	Xylitol	Arabinose $\xrightarrow{EC:5.3.1.4}$ Ribulose $\xrightarrow{EC:5.3.1.20}$ Ribose $\xrightarrow{EC:2.7.1.15}$ Ribose-5P $\xrightarrow{EC:2.2.1.2, EC:2.2.1.1}$ Xylulose-5P $\xrightarrow{EC:2.7.1.17}$ Xylulose $\xrightarrow{EC:1.1.1.15, EC:1.1.1.10, EC:1.1.1.9}$ Xylitol
Galactose	Leloir pathway	Galactose $\xrightarrow{EC:2.7.1.6}$ Galactose-1P $\xrightarrow{EC:2.7.7.12}$ Glucose-1P $\xrightarrow{EC:5.4.2.5, EC:5.4.2.2}$ Glucose-6P
	Ribitol	Galactose $\xrightarrow{\text{Leloir pathway}}$ Glucose-6P $\xrightarrow{M00006}$ Ribulose-5P $\xrightarrow{EC:2.7.1.16}$ Ribulose $\xrightarrow{EC:1.1.1.56}$ Ribitol
	Arabitol	Galactose $\xrightarrow{\text{Leloir pathway}}$ Glucose-6P $\xrightarrow{M00006}$ Ribulose-5P $\xrightarrow{EC:2.7.1.16}$ Ribulose $\xrightarrow{EC:1.1.1.13}$ Arabitol
	Erythritol	Galactose $\xrightarrow{\text{Leloir pathway}}$ Glucose-6P $\xrightarrow{EC:5.3.1.9}$ Fructose-6P $\xrightarrow{EC:2.2.1.1}$ Erythrose-4P $\xrightarrow{EC:2.7.1.27}$ Erythritol
	Mannitol	Galactose $\xrightarrow{\text{Leloir pathway}}$ Glucose-6P $\xrightarrow{EC:5.3.1.9}$ Fructose-6P $\xrightarrow{EC:1.1.1.17}$ Mannitol-1P $\xrightarrow{EC:3.1.3.22}$ Mannitol
	Xylitol	Galactose $\xrightarrow{\text{Leloir pathway}}$ Glucose-6P $\xrightarrow{M00006}$ Ribulose-5P $\xrightarrow{EC:5.1.3.1}$ Xylulose-5P $\xrightarrow{EC:2.7.1.17}$ Xylulose $\xrightarrow{EC:1.1.1.15, EC:1.1.1.10, EC:1.1.1.9}$ Xylitol

또한 이들은 단순히 연결된 것이 아니라 실제로 당알코올의 축적과 당의 결핍이 있다는 것을 확인해보았다. Figure 8b 에 따르면 Xylitol 이 arabinose 의 결핍에 관여하는 정도만 아주 낮은 값을 보였으며, 나머지 당알코올의 축적과 당의 결핍은 아주 높은 상관관계를 가지고 있음을 확인하였다.

그러나 당알코올의 축적이 실제로 당의 결핍에 영향을 끼쳤는지는 상호관계가 있다는 것과는 다르다. 따라서 당알코올의 축적이 group variable (정상 군과 감염 군)과의 교호작용 (interaction effect)를 가져 ANM substrates 의 감소에 영향을 미친다는 것을 multiple linear regression model 을 이용하여 증명하였다. 수식과 그 통계 모델의 적용 과정은 아래와 같다.

$$Y^i = \beta_0^{ij} + \beta_1^{ij}G + \beta_2^{ij}X^j + \beta_3^{ij}(G \times X^j)$$

$$i = 1,2,3, j = 1,2,3,4,5$$

where Y^i is the ANM supply, with i (1:fructose; 2:arabinose; 3:galactose) as the dependent variable, and G and X^j are independent variables denoting the group (0: control; 1: parasitism) and the accumulated sugar alcohol j (1:ribitol; 2:arabitol; 3:erythritol; 4:mannitol; 5:xylitol). Overall, 45 multiple linear regression models and interaction effect coefficients, β_3^{ij} s, between the group variable and metabolite j to metabolite i were included in the model (Table 1). The hypothesis regarding the existence of an interaction effect is $H_0 : \beta_3^{ij} = 0$. Crossing of regression lines (regression validated by the adjusted R square in Table 1) on the scatter plots crossed owing to a slope difference suggested that the group variable and the sugar alcohols had an interaction effect on the expression of dependent variables

이를 이용하여 계산한 교호작용의 수치는 각 당알코올이 각 ANM substrates 에 관여하는 정도에 대해서 유의성이 있는지를 보여준다 (Table 5). 교호작용의 방향이 -값을 갖는 경우 ANM supply 가 당알코올의 축적에 의하여 감소한다는 것을 뜻하며, 그 정도는 adjusted R square, 그리고 그것이 유의성있음을 p-value 로써 설명한다. Galactose 의 감소는 교호작용의 방향은 맞으나, 유의성있는 영향을 끼치는 것이 총 15 개 (각 종별로 5 개의

당알코올)중 2 개밖에 없어 galactose 의 감소는 당알코올의 영향을 적게 받는 것을 알 수 있다. 그러나 다른 ANM supply 의 감소에는 CP, PS, CS 에서 유의성 있게 효과를 갖는 경우가 다수였다. 결론적으로 당알코올의 축적은 ANM supply 의 감소에 영향이 있다는 것을 확인하였다.

Table 5. Direction, adjusted R square and p-values of the interaction effect between sugars (fructose, arabinose and galactose; dependent variables) and sugar alcohols (ribitol, arabitol, erythritol, mannitol and xylitol; independent variables). Adjusted R square indicates a degree of interaction between them, and the p-value indicates the reliability of the effect.

Species	Metabolite	Fructose			Arabinose			Galactose		
		Effect direction	Adjusted R square	p-value	Effect direction	Adjusted R square	p-value	Effect direction	Adjusted R square	p-value
CP	Ribitol	+	0.534	0.346	+	0.853	0.184	-	0.584	0.854
	Arabitol	-	0.924	<0.001	-	0.983	<0.001	-	0.787	0.102
	Erythritol	-	0.497	0.421	+	0.837	0.436	-	0.594	0.872
	Mannitol	-	0.888	<0.001	-	0.953	<0.001	-	0.775	0.070
	Xylitol	-	0.914	<0.001	-	0.982	<0.001	-	0.760	0.020
PS	Ribitol	+	0.577	0.497	-	0.743	0.277	-	0.797	0.868
	Arabitol	+	0.625	0.039	-	0.567	0.014	-	0.769	0.528
	Erythritol	+	0.520	0.695	+	0.789	0.868	-	0.892	0.637
	Mannitol	-	0.532	0.952	+	0.548	0.743	-	0.769	0.781
	Xylitol	+	0.640	0.032	-	0.888	0.150	-	0.852	0.237
CS	Ribitol	+	0.909	0.980	-	0.816	0.158	+	0.563	0.144
	Arabitol	-	0.967	0.028	-	0.925	0.028	-	0.841	0.120
	Erythritol	-	0.953	0.067	-	0.893	0.002	-	0.771	0.241
	Mannitol	-	0.969	0.012	-	0.866	0.018	-	0.787	0.348
	Xylitol	-	0.981	0.180	-	0.946	0.041	-	0.811	0.003

기존의 연구에서 plant-microbe interaction 시 fructose 가 mannitol 로 변환되어 균이 사용한다는 것과[22, 25] glucose, fructose 가 hexose transporter 에 의해 숙주에서 균으로 수송된다는 것이 밝혀져 있다 [26-29]. 그러나 mannitol 외의 다른 당알코올의 축적이 몇 C6 당의 감소를 야기한다는 보고는 없었으며, 이 모든 당알코올의 축적이 ANM pathway 의 기질로써 사용되는 당의 결핍을 가져온다는 것은 밝혀진 바 없었다 (Figure 9).

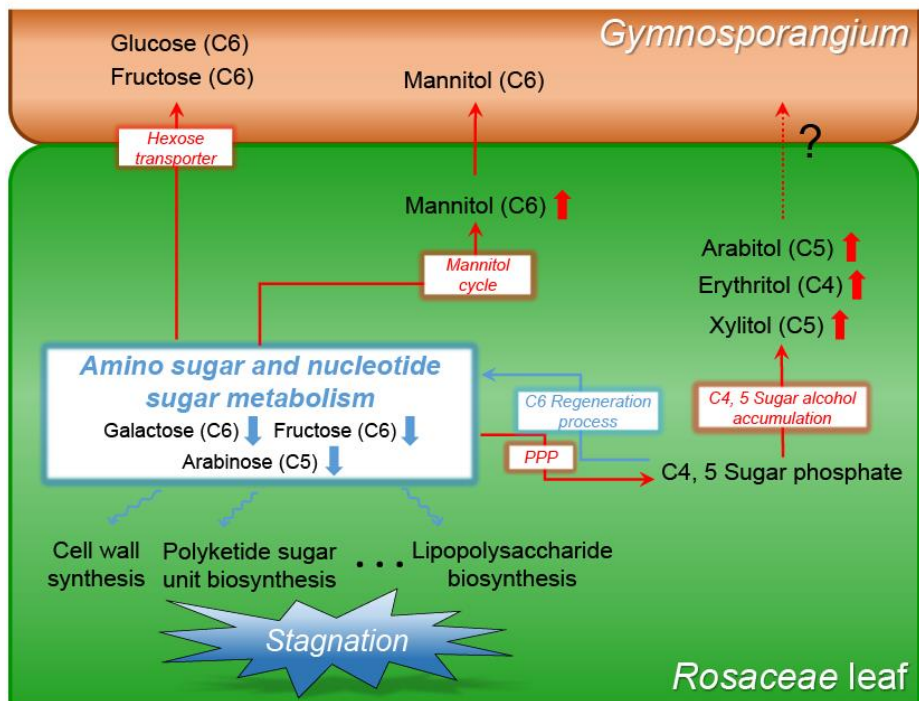


Figure 9. Abnormal accumulation of C4 and C5 sugar alcohols leads to metabolic deficiency in the ANM pathway. Red arrows indicate increases in metabolites or pathway activity; blue arrows indicate the opposite effect.

이러한 ANM pathway 의 감소가 중요한 이유는, 이 기전의 products 들은 균의 cell wall synthesis 에서 sugar unit 의 공급자 역할을 하는 nucleotide 의 phosphate form 이기 때문이다 [30-35]. ANM pathway 에 결핍이 올 경우 cell wall 의 유지가 힘들기

때문에 균이 숙주에 접촉하였을 때 이를 뚫고 haustoria 를 생성하고, 이를 다시 막으려는 cell wall repairing 이 진행되지 않을 수 있다. 따라서 적성병에 균체를 생성하는 증상이 해당 대사기전의 결핍이 연관되어 있을 수 있다는 가능성을 확인할 수 있다. 해당 연구결과는 본 저자가 Scientific reports 에 게재한 ‘Metabolic response induced by parasitic plant–fungus interactions hinder amino sugar and nucleotide sugar metabolism in the host’ 를 참조하여 작성하였다 [36].

IV. 결론

이 연구는 적성병균인 *Gymnosporangium asiaticum*이 숙주인 장미과 식물을 감염시켰을 때 유도되는 대사기전의 변화가 질병의 증상에 관련이 있다는 것을 밝혔다. 구체적으로는 장미과 식물에 비정상적인 당알코올 축적 유도를 통하여 Aminosugar and nucleotide sugar metabolism의 substrate의 결핍을 가져온다는 것을 확인하였고, 이것이 숙주의 세포벽 생성 등을 억제하여 균주의 침입을 가능하게 했다는 것이다. 이러한 결론을 위해 우리는 새로이 상관관계 네트워크 분석을 사용하였다. 대사기전의 enrichment부터 pathway mapping, 그리고 그 대사체들간의 상호관계를 연구하기 위한 교호작용 확인 등의 네트워크 분석법은 기존의 대사체학에서 많이 활용되지 않았다. 따라서 이 논문에서 대사체학의 새로운 방식을 열었으며, 이러한 접근법은 향후 다양한 대사체학 연구, 특히 병원성 대사기전 변화에 대한 특정 대사체 사이의 영향력 여부를 연구하는 데 있어서 유용한 도구가 될 것이다.

V. 참고문헌

1. Mendgen, K. and M. Hahn, *Plant infection and the establishment of fungal biotrophy*. Trends in Plant Science, 2002. **7**(8): p. 352–356.
2. Spanu, P.D., et al., *Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism*. Science, 2010. **330**(6010): p. 1543–1546.
3. Lyu, X., et al., *Comparative genomic and transcriptional analyses of the carbohydrate-active enzymes and secretomes of phytopathogenic fungi reveal their significant roles during infection and development*. Scientific Reports, 2015. **5**: p. 15565.
4. Misra, R.C., et al., *A thaumatin-like protein of *Ocimum basilicum* confers tolerance to fungal pathogen and abiotic stress in transgenic *Arabidopsis**. Scientific Reports, 2016. **6**: p. 25340.
5. Udvardi, M. and P.S. Poole, *Transport and Metabolism in Legume-Rhizobia Symbioses*. Annual Review of Plant Biology, 2013. **64**(1): p. 781–805.
6. Hofmann, J., et al., *Metabolic profiling reveals local and systemic responses of host plants to nematode parasitism*. The Plant Journal, 2010. **62**(6): p. 1058–1071.
7. Jeelani, G. and T. Nozaki, *Metabolomic analysis of *Entamoeba*: applications and implications*. Current Opinion in Microbiology, 2014. **20**(0): p. 118–124.
8. Yun, H.-Y., S.-K. Lee, and K.-J. Lee, *Identification of Aecial Host Ranges of Four Korean Gymnosporangium Species Based on the Artificial Inoculation with Teliospores Obtained from Various Forms of Telia*. The Plant Pathology Journal, 2005. **21**(4): p. 310–316.
9. Zhou, K., et al., *Metabonomics reveals metabolite changes in biliary atresia infants*. Journal of Proteome research, 2015. **14**(6): p. 2569–74.
10. Zhao, Y., et al., *A metabolomics study delineating geographical location-associated primary metabolic changes in the leaves of growing tobacco plants by GC-MS and CE-MS*. Scientific Reports, 2015. **5**: p. 16346.
11. Kind, T., et al., *A comprehensive urinary metabolomic approach for identifying kidney cancer*. Analytical Biochemistry, 2007. **363**(2): p. 185–195.
12. Katajamaa, M. and M. Orešič, *Data processing for mass spectrometry-based metabolomics*. Journal of Chromatography A, 2007. **1158**(1–2): p. 318–328.
13. Sreekumar, A., et al., *Metabolomic profiles delineate potential*

- role for sarcosine in prostate cancer progression. *Nature*, 2009. **457**(7231): p. 910–914.
14. Czauderna, T., C. Klukas, and F. Schreiber, *Editing, validating and translating of SBGN maps*. *Bioinformatics*, 2010. **26**(18): p. 2340–2341.
 15. Chagoyen, M. and F. Pazos, *MBRole: enrichment analysis of metabolomic data*. *Bioinformatics*, 2011. **27**(5): p. 730–1.
 16. Harrington, T.C., D. McNew, and H.Y. Yun, *Bur oak blight, a new disease on Quercus macrocarpa caused by Tubakia iowensis sp. nov.* *Mycologia*, 2012. **104**(1): p. 79–92.
 17. Van der Auwera, G., S. Chapelle, and R. De Wlchter, *Structure of the large ribosomal subunit RNA of Phytophthora megasperma, and phylogeny of the oomycetes*. *FEBS Letters*, 1994. **338**(2): p. 133–136.
 18. Yun, H.Y., et al., *The rust fungus Gymnosporangium in Korea including two new species, G. monticola and G. unicorne*. *Mycologia*, 2009. **101**(6): p. 790–809.
 19. Jeon, Y.-S., et al., *jPHYDIT: a JAVA-based integrated environment for molecular phylogeny of ribosomal RNA sequences*. *Bioinformatics*, 2005. **21**(14): p. 3171–3173.
 20. Bowne, J.B., et al., *Drought Responses of Leaf Tissues from Wheat Cultivars of Differing Drought Tolerance at the Metabolite Level*. *Molecular Plant*, 2012. **5**(2): p. 418–429.
 21. Huck, J.H.J., et al., *Profiling of Pentose Phosphate Pathway Intermediates in Blood Spots by Tandem Mass Spectrometry: Application to Transaldolase Deficiency*. *Clinical Chemistry*, 2003. **49**(8): p. 1375–1380.
 22. Solomon, P.S., O.D. Waters, and R.P. Oliver, *Decoding the mannitol enigma in filamentous fungi*. *Trends in microbiology*, 2007. **15**(6): p. 257–62.
 23. Najjar, V.A. and M.E. Pullman, *The Occurrence of a Group Transfer Involving Enzyme (phosphoglucomutase) and Substrate*. *Science*, 1954. **119**(3097): p. 631–634.
 24. Holden, H.M., I. Rayment, and J.B. Thoden, *Structure and Function of Enzymes of the Leloir Pathway for Galactose Metabolism*. *Journal of Biological Chemistry*, 2003. **278**(45): p. 43885–43888.
 25. Voegelé, R.T., et al., *Possible roles for mannitol and mannitol dehydrogenase in the biotrophic plant pathogen Uromyces fabae*. *Plant Physiology*, 2005. **137**(1): p. 190–8.
 26. Schulze-Lefert, P. and R. Panstruga, *Establishment of biotrophy by parasitic fungi and reprogramming of host cells for disease resistance*. *Annual Review of Phytopathology*, 2003. **41**(1): p. 641–667.
 27. Bonfante, P. and A. Genre, *Mechanisms underlying beneficial*

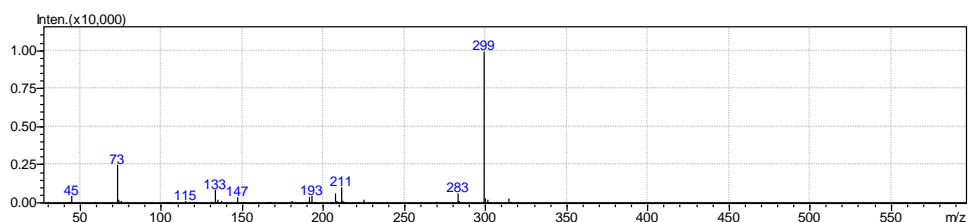
- plant–fungus interactions in mycorrhizal symbiosis*. Nature Communications, 2010. **1**: p. 48.
28. Hahn, M. and K. Mendgen, *Signal and nutrient exchange at biotrophic plant–fungus interfaces*. Current Opinion in Plant Biology, 2001. **4**(4): p. 322–327.
 29. Voegelé, R.T., et al., *The role of haustoria in sugar supply during infection of broad bean by the rust fungus Uromyces fabae*. Proceedings of the National Academy of Sciences, 2001. **98**(14): p. 8133–8138.
 30. Clay, N.K., et al., *Glucosinolate Metabolites Required for an Arabidopsis Innate Immune Response*. Science, 2009. **323**(5910): p. 95–101.
 31. Luna, E., et al., *Callose Deposition: A Multifaceted Plant Defense Response*. Molecular Plant–Microbe Interactions, 2010. **24**(2): p. 183–193.
 32. Schlupmann, H., A. Bacic, and S.M. Read, *Uridine Diphosphate Glucose Metabolism and Callose Synthesis in Cultured Pollen Tubes of Nicotiana glauca Link et Otto*. Plant Physiology, 1994. **105**(2): p. 659–670.
 33. Wang, Q., et al., *Identification of a UDP–glucose pyrophosphorylase from cotton (Gossypium hirsutum L.) involved in cellulose biosynthesis in Arabidopsis thaliana*. Plant Cell Reports, 2011. **30**(7): p. 1303–1312.
 34. Joshi-Estanyol, M. and P. Puigdomènech, *Plant cell wall glycoproteins and their genes*. Plant Physiology and Biochemistry, 2000. **38**(1–2): p. 97–108.
 35. Burton, R.A., M.J. Gidley, and G.B. Fincher, *Heterogeneity in the chemistry, structure and function of plant cell walls*. Nature Chemical Biology, 2010. **6**(10): p. 724–32.
 36. Lee, D.-K., et al., *Metabolic response induced by parasitic plant–fungus interactions hinder amino sugar and nucleotide sugar metabolism in the host*. Scientific reports, 2016. **6**(37434).

VI. 부록

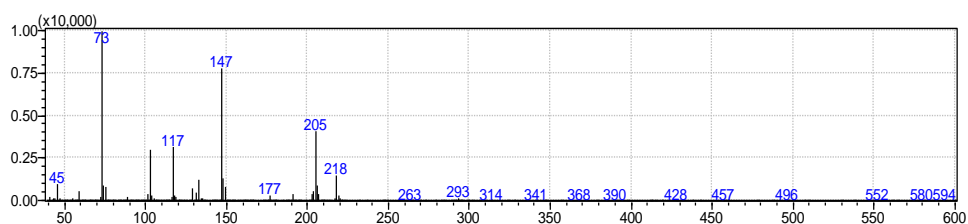
1. GC-MS spectrum of 34 metabolites

Mass spectra of derivatized metabolites. The list of metabolites were as follows: phosphoric acid (a), glycerol (b), succinic acid (c), glyceric acid (d), fumaric acid (e), serine (f), threonine (g), citramalic acid (h), malic acid (i), erythritol (j), aspartic acid (k), erythronic acid (l), 3-hydroxy-3-methylpentanedioic acid (m), xylonic acid (n), xylose (o), arabinose (p), ribose (q), xylitol (r), arabitol (s), ribitol (t), arabinofuranose (u), ribonic acid (v), idonic acid (w), citric acid (x), sorbose (y), mannose (z), fructose (aa), galactose (ab), mannitol (ac), sorbitol (ad), gluconic acid (ae), inositol (af), glucose (ah), linoleic acid (ah), oleic acid (ai), stearic acid (aj), sucrose (ak), maltose (al) and melibiose (am)

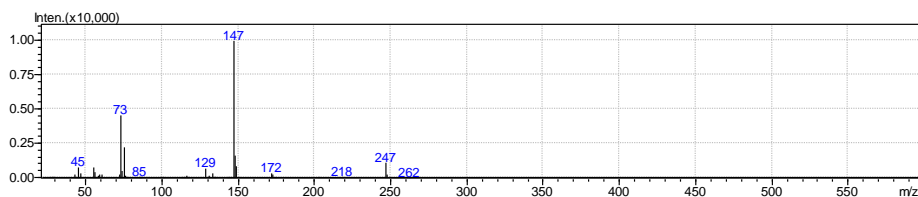
(a)



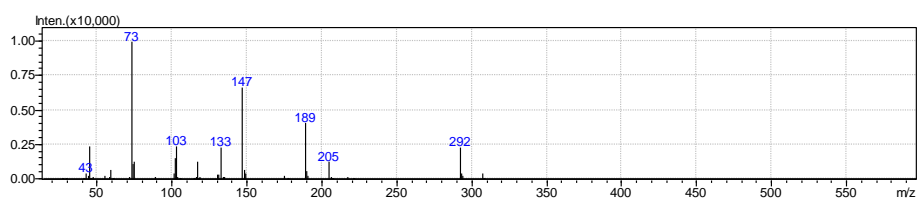
(b)



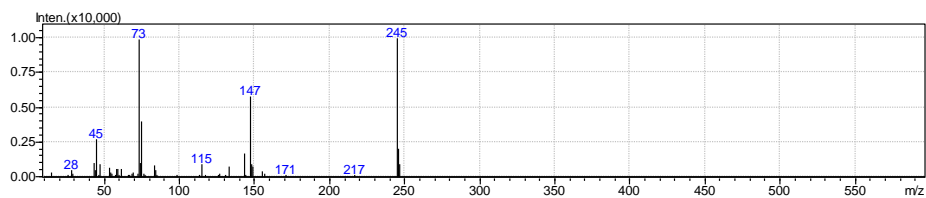
(c)



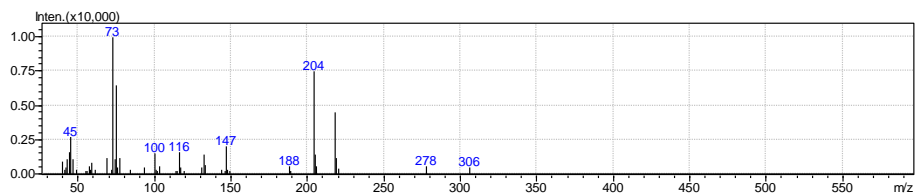
(d)



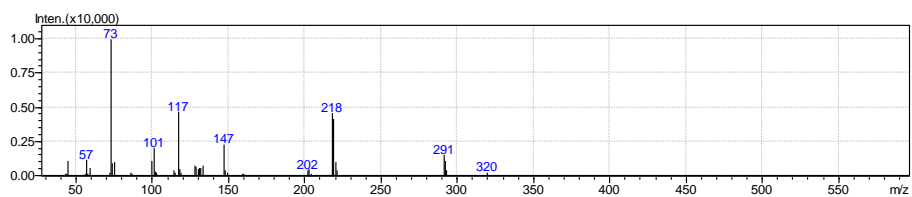
(e)



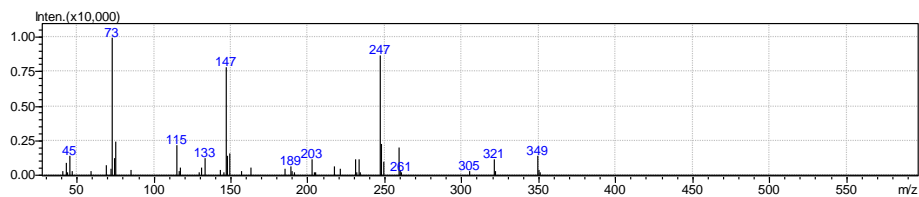
(f)



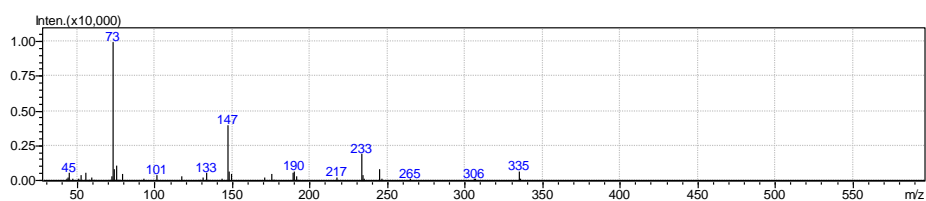
(g)



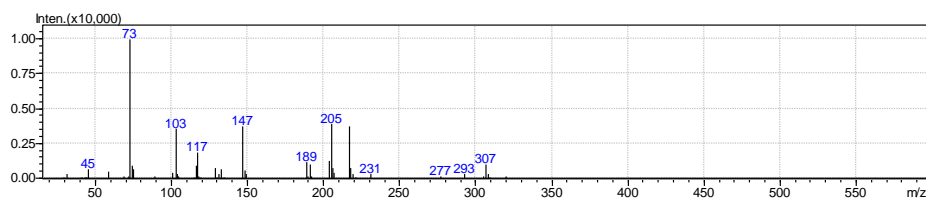
(h)



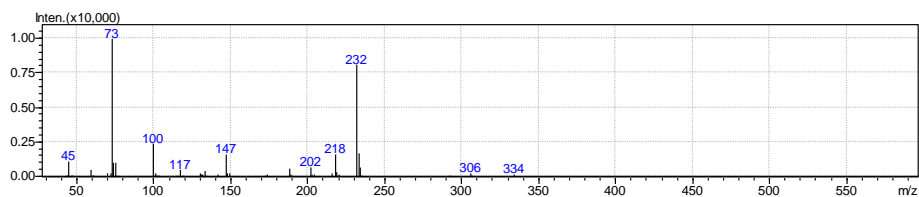
(i)



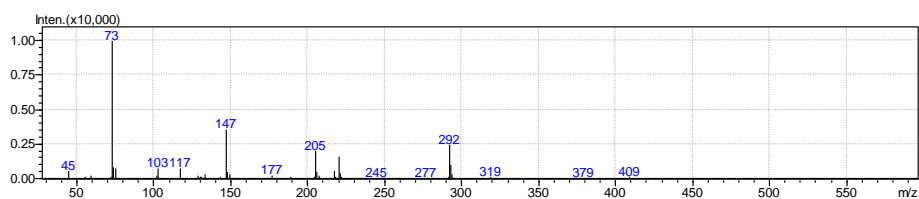
(j)



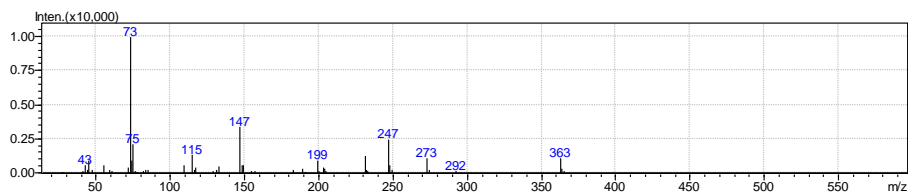
(k)



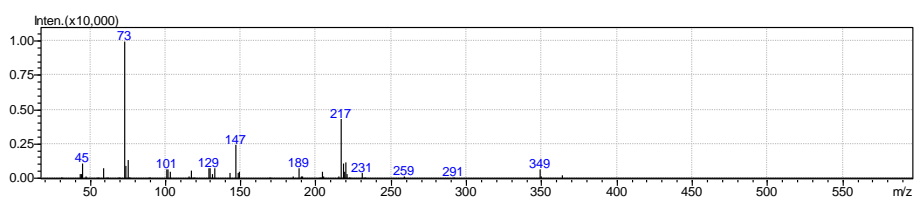
(l)



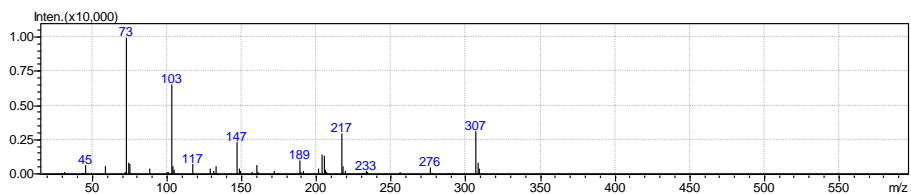
(m)



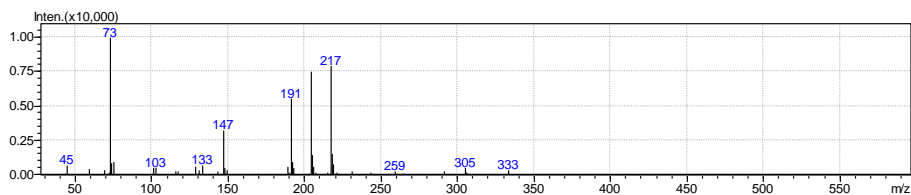
(n)



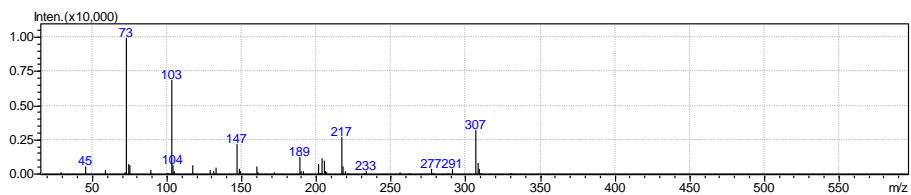
(o)



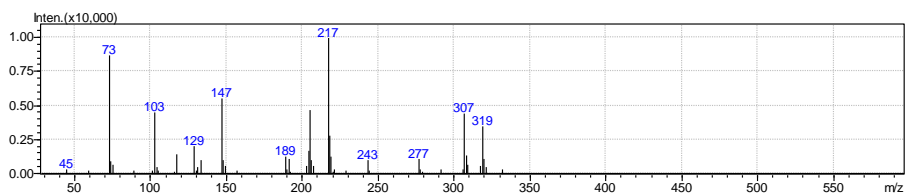
(p)



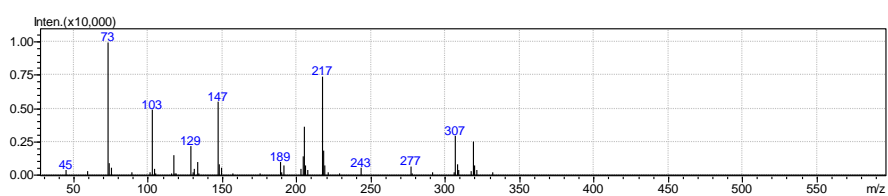
(q)



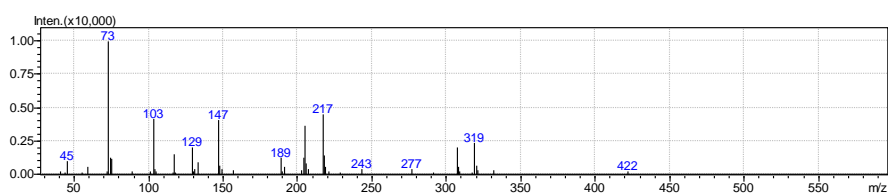
(r)



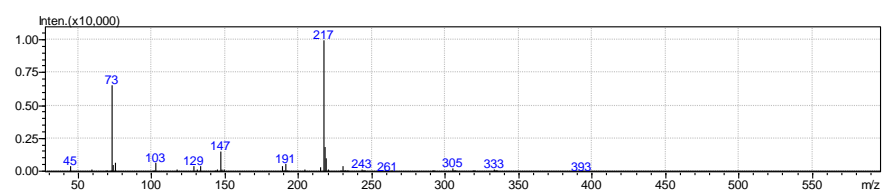
(s)



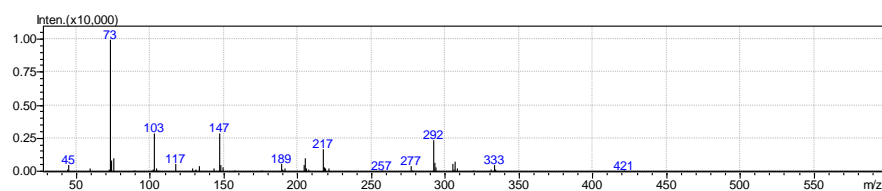
(t)



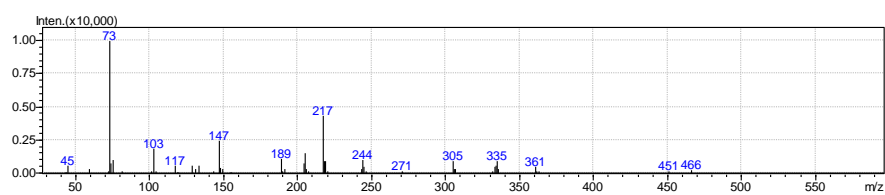
(u)



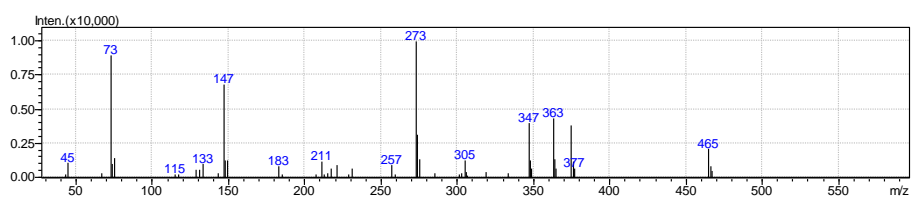
(v)



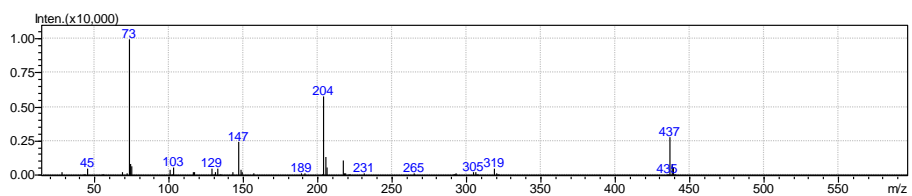
(w)



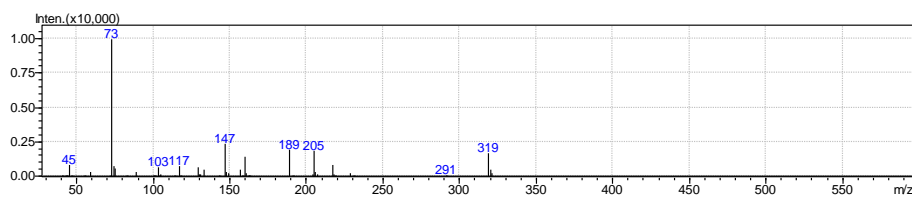
(x)



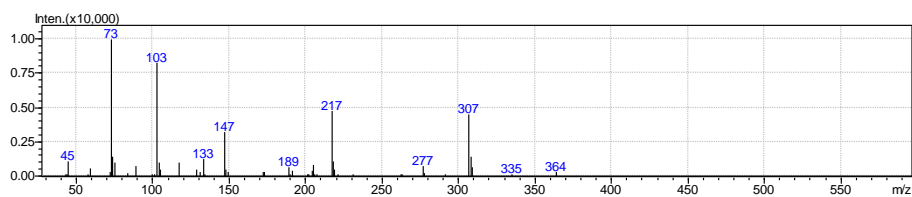
(y)



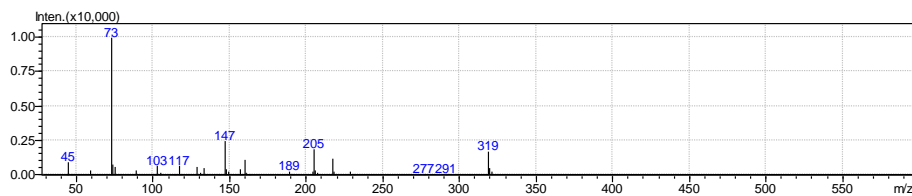
(z)



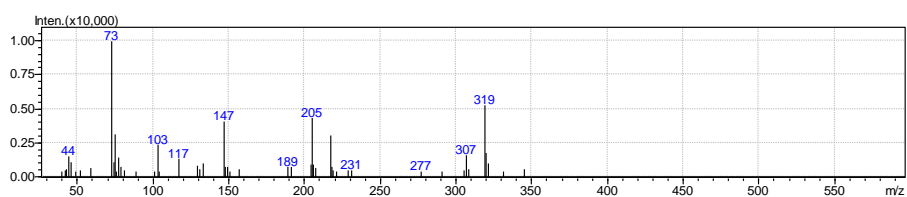
(aa)



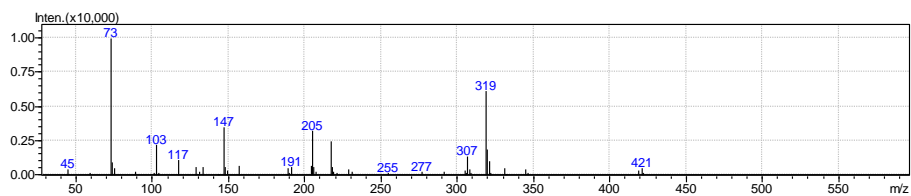
(ab)



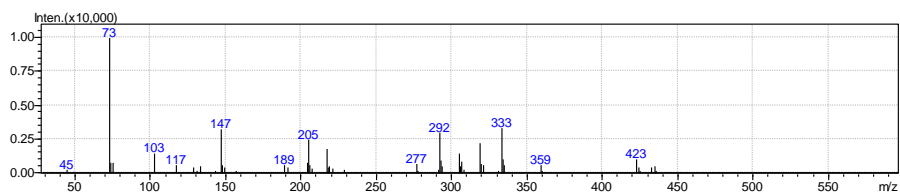
(ac)



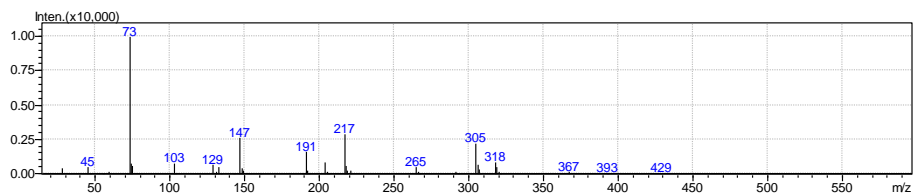
(ad)



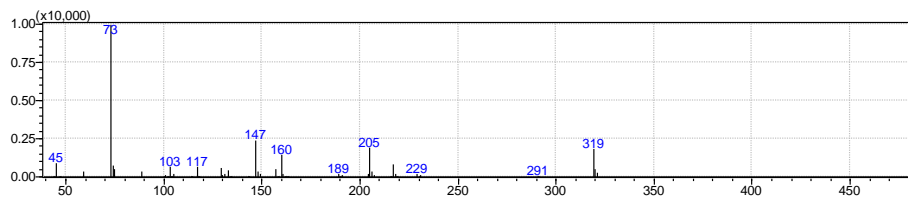
(ae)



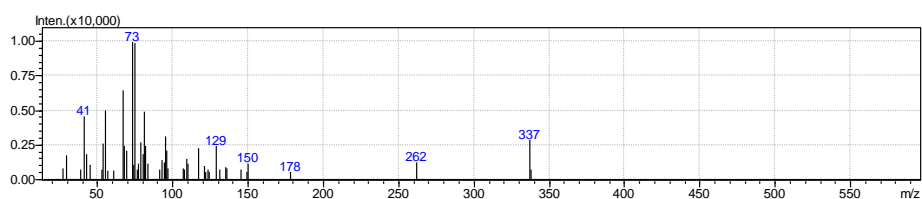
(af)



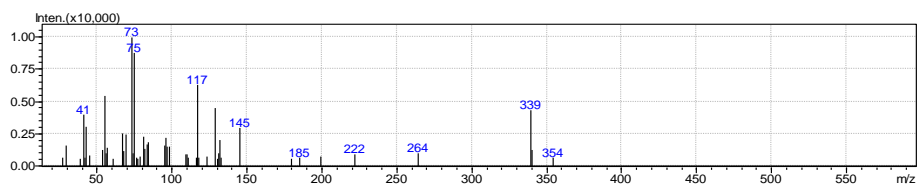
(ag)



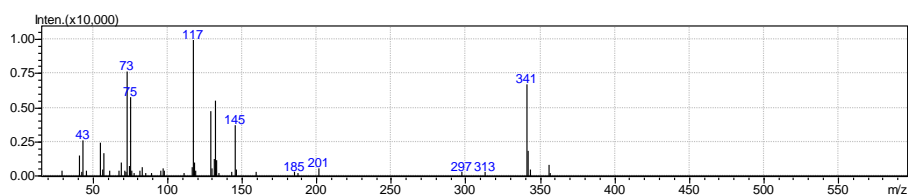
(ah)



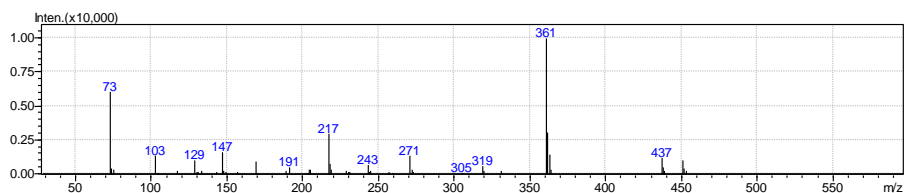
(ai)



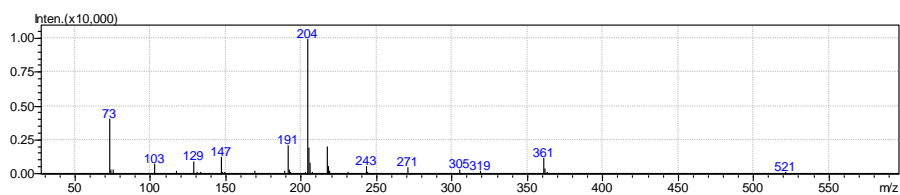
(aj)



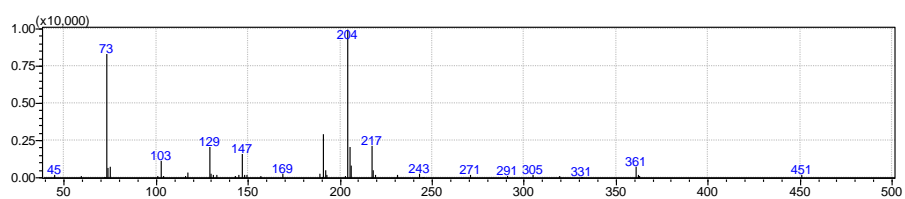
(ak)



(al)



(am)



2. GC-MS 기반 대사체학

대사체학을 적용하는데 있어 spectroscopy의 선정이 매우 중요하다. 각 분광기가 해석할 수 있는 물성 또는 화학적 특성에 따라서 실험자가 관찰하고자 하는 대사체를 검출할 수 있는지에 대한 여부가 결정되기 때문이다. 대사체를 검출하는데 흔히 사용되는 분광기로써 근적외선 분광기 (Near-infrared spectroscopy, NIR), 핵자기공명분광기 (Nuclear magnetic resonance spectroscopy; NMR), 질량분석기 (Mass spectrometry; MS) 가 있다.

NIR의 경우는 생물체에 존재하는 수만가지 대사체를 작용기에 기인한 분류를 통하여 검출할 수 있게 해준다. 가장 큰 강점은, 데이터를 획득하는 시간이 굉장히 빠르다는 점, 그리고 시료에 특별한 전처리를 거치지 않은 상태에서 데이터 획득이 가능하다는 점이다. 그러나 NIR의 경우는 각 분자를 구분해내지 못하며, 작용기로 분자들을 묶어서 한꺼번에 정량하는 기기이기에, 구체적인 대사기전예의 적용은 불가하다. 따라서 특정 실험군에 따른 대사기전 변화를 관찰하는데는 부적합하며, 대사체의 양적 변화를 기반으로 한 판별에 적합한 기기이다.

NMR의 경우는 분자 내 각각의 원소들이 가지고 있는 핵의 스핀을 이용하여 검출한다. 외부의 자기장이 존재하지 않는 경우 핵 스핀이 무작위의 방향성을 가지고 있다. 이때 NMR을 이용하여 유도자기장을 전사할 경우 핵 스핀이 정렬되고, 이때 특정 주파수를 외부에서 제공하게 되면 각 원소의 핵 스핀의 강도에 따라서 다른 주파수를 흡수하게 되며, 이러한 원리로 대사체의 원소를 검출할 수 있다. 자기장을 걸었을 때 에너지 차이가 존재하기 위해서는 홀수분자량을 갖는 원소를 가져야 하며, 그렇기에 ^1H (natural abundance: 99.98%), ^{13}C (natural

abundance: 1.1%) 를 주요 구성성분으로 하는 탄화수소, 즉 유기물들을 검출할 수 있으며, 또한 ^{31}P (natural abundance: 100%)등의 다양한 무기물이 결합된 형태의 대사체를 검출하는데 사용할 수 있다. NMR을 이용하여 대사체를 검출할 때 사용할 수 있는 주요 원소들은 아래와 같다.

Isotope ^[5]	Mag Moment (μ , nm) ^[4]	Nuclear Spin ^[4]	Natural Abundance (%) ^[4]
^1H	2.79284734(3)	1/2	~100
^2H	0.857438228(9)	1	0.015
^3H	2.97896244(4)	1/2	0
^{10}B	1.80064478(6)	3	19.9
^{11}B	2.6886489	3/2	80.1
^{13}C	0.7024118(14)	1/2	1.1
^{14}N	0.40376100(6)	1	99.6
^{15}N	-0.28318884(5)	1/2	0.37
^{17}O	-1.89379(9)	5/2	0.04
^{19}F	2.628868(8)	1/2	~100
^{31}P	1.13160(3)	1/2	~100

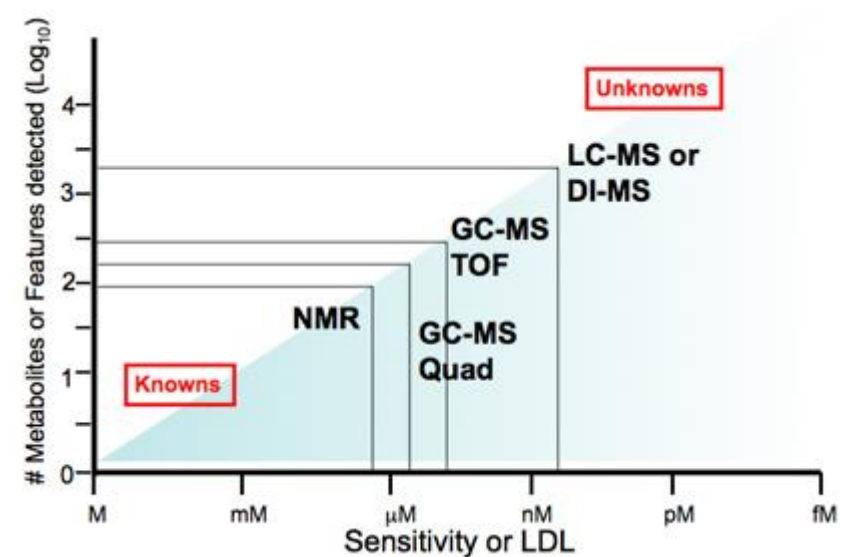
http://en.wikipedia.org/wiki/Nitrogen-15_nuclear_magnetic_resonance_spectroscopy

따라서 검출 가능한 분자의 범위가 굉장히 넓은 편이며, 특히 미지의 물질을 검출하였을 때 구조의 분석이 가능하여 특히나 천연물의 대사체 구조 분석에 많이 사용되고 있다.

그러나 NMR을 이용하여 생체 시료에 포함된 대사체를 분석하는 데에는 몇가지 단점이 있다. 일단 각 대사체에 포함된 모든 원소의 peak가 한 스펙트럼에 나타나 overlay되는 경우가 허다하며, 따라서 각 대사체 고유의 chemical shift에서 단일 peak로 검출되지 않는 경우 대사체의 profiling이 어렵다. 또한 낮은 sensitivity를 갖기에 극미량 함유된 대사체는 검출이 되지 않는다는 단점이 있다. 따라서 실제 생체시료

에서 대사체 profiling을 할 때 검출되는 대사체의 개수는 10~30개에 국한된다.

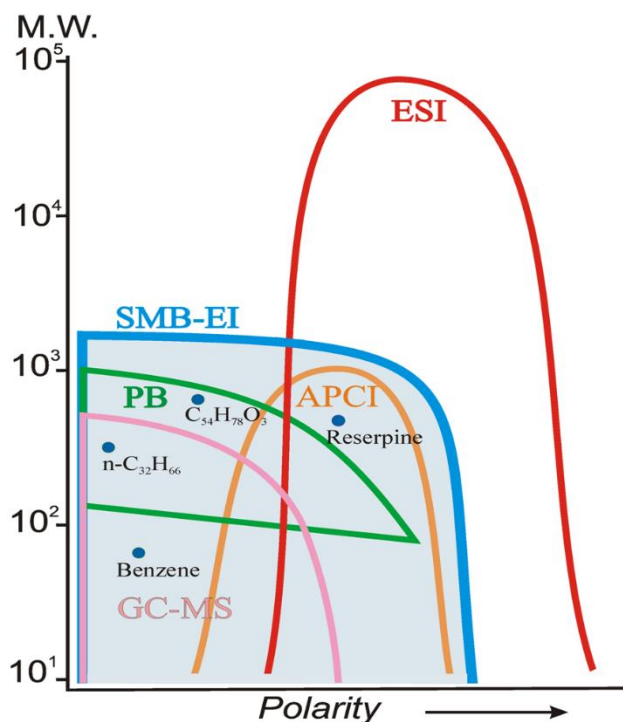
위의 두가지 기기에 비해 가장 최근에 많이 사용되는 것이 질량분석기에 기반한 대사체학이다. 분자의 이온화가 비교적 쉽게 가능하여 일찍부터 시작한 기체크로마토그래피-질량분석기 (Gas chromatography-mass spectroscopy, GC-MS)와, 이온화방식의 개발이 조금 늦어 최근에서야 급부상한 액체크로마토그래피-질량분석기 (Liquid chromatography-mass spectrometry)가 있다. 이 두가지 기기를 모두 포함하여 검출할 수 있는 대사체의 종류는 약 5000종류에 가까우며, 이는 위의 나머지 기기들에 비하면 굉장히 많은 종류의 대사체를 검출할 수 있다는 점에서 최근에 각광받고 있다.



이들은 Chromatography수준에서의 분자의 분리와 질량분석기 차원에서의 분리능을 함께 고려하여 다양한 분자들 각각의 정량을 확인할 수 있다는 특성을 가진다. GC-MS와 LC-MS 모두 물질의 특성에 따라서 다양한 성분을 골고루 분리할 수 있는 다양한 종류의 컬럼을 선택할 수 있어 이로부터 한번에 100여종의 대사체 profiling이 가능하다.

또한 위의 기기들에 비해 압도적인 sensitivity를 가지기에 앞의 기기들에서 검출되지 않았던 극미량의 대사체들이 검출될 수 있다는 장점을 가지고 있다.

그런데 GC-MS와 LC-MS도 검출하고자 하는 물질에 따라서 나누어 사용이 된다. LC-MS의 경우는 이온화 방식에 따라 조금씩은 차이가 나지만 GC-MS에 비해 더 넓은 검출 범위를 갖는다. 특히 검출이 가능한 물질을 구분하는 것이 질량분석기가 아닌, 앞의 분리를 담당하는 액체크로마토그래피와 기체크로마토그래피가 주로 검출가능한 물질의 종류를 결정한다.



<http://blog.avivanalytical.com/2013/06/what-can-be-improved-in-gc-ms-when.html>

LC-MS는 이온분무화 방식 (Electrospray ionization)을 사용하여, 주로 분자량이 커서 휘발되지 못하는 물질의 분석에 활용되며, 다양한 컬럼에 따라 분리되는 물질을 다르지만, 주로 semi-polar metabolites인

secondary metabolites가 주로 검출된다. 따라서 주로 lipid 및 고분자물질의 분석에 LC-MS가 사용된다. 그리고 TQ (Triple-quadrupole), Q-TOF (Quadrupole-Time of flight)와 같은 detector의 발전에 힘입어, HILIC (Hydrophilic interaction chromatography) separation에 기반한 polar compounds의 검출도 가능해지고 있다. 그러나 아직은 Chromatographic resolution이 GC-MS에 비해 낮은 편이기에, Polar metabolites이며, 상대적으로 저분자 (MW 500 이하)인 경우는 GC-MS를 이용하여 검출하는 것이 아직은 더 우수한 성능을 보인다. 따라서 현재 secondary metabolites 및 non-polar한 lipid의 분석에는 LC-MS를 주로 활용하며, amino acid, sugar와 같은 polar compounds는 GC-MS를 사용하여 검출한다.

GC-MS의 경우는 위에서 설명한 바대로, polar compound의 분석에 주로 사용된다. GC-MS는 Chemical ionization (CI), Electron impact (EI) ionization 방식을 사용하며, 근래에는 EI에 기반한 데이터베이스가 형성되어, 대사체학에는 주로 EI방식을 사용한다. GC-MS의 검출에서 가장 큰 한계점은 Injection temperature에서 기화가 되어야 한다는 점인데, polar metabolites들은 원래의 형태에서는 대체로 기화가 되지 않는다. 따라서 이러한 단점을 보완하기 위하여 각 polar functional group을 blocking하여 휘발성을 부여하는 유도체화 방식을 주로 이용하고 있다. 유도체화에는 Silylation, Acylation, Alkylation, Esterification 등이 사용되고 있으며, 대사체 profiling에는 주로 methoxyamination+trimethylsilylation 의 multistep derivatization이 사용된다.

이 섹션에서는 해당 학위논문에서 대사체 profiling하는 방법에 대해서 구체적으로 서술하고자 하며, GC-MS에 기반한 대사체학을 할 경우 샘플 준비부터 데이터의 획득까지 필요한 모든 부분에 대해서 서술하고자 한다.

1) 대사체 추출

대사체의 추출에는 주로 Alcohol류의 solvent, 또는 이들과 Water 또는 Chloroform등의 용매를 섞은 용액을 주로 사용한다. 추출은 주로 Sonication을 하며, 단 식물 및 동물이 세포막 및 세포벽의 유무 등이 다르기 때문에 추출 방식을 달리한다.

일단 추출에 앞서 샘플을 collection 할 때 metabolism quenching이 선행되어야 한다. quenching에는 -70°C 이하에서 보존하는 방식이 있으며, 가장 완벽하게는 액체질소에서 보존하는 방식이 있다. 식물과 동물 모두 특정 실험 군의 처리가 된 후 변화를 막기 위해 metabolism quenching이 되어야 한다. 또한 metabolism quenching과 추출을 동시에 진행하는 방식도 있으며, 저온의 추출용매를 ammonium bicarbonate와 같은 buffer와 함께 가하여 quenching 및 추출을 동시에 진행한다.

식물의 대사체 추출은 아주 기본적으로 용매를 가하여 추출하는 방식을 사용한다. Alcohol류 용매를 이용한 추출로부터 Primary metabolites와 Secondary metabolites를 모두 취할 수 있다. 식물샘플에서는 GC-MS에서 휘발되지 못하며, 아직 밝혀지지 않은 미지의 물질이 다량함유되어 있어, 추출 후 전처리가 더욱 중요하다. 1차적으로 $0.2\text{ }\mu\text{m}$ filter를 이용하여 용해되지 않은 물질들을 제거해주어야 하며, 그 후 secondary metabolites 들을 제거하기 위한 SPE sampling을 진행해주어야 한다. 단, liner에 의한 기기 자체적인 filtering 등이 있어 SPE sampling은 생략할 수 있으나, 기기 사용 단계에서 지속적인 관리가 필요하다. 그 후 용매의 구성에 맞게 Nitrogen purge, SpeedVac, freeze-drying 을 이용하여 용매를 휘발시켜 준다.

동물 시료의 경우는 식물과 다르게 세포를 파쇄하는 작업이 필요하다. 추출용매는 식물에서의 대사체 추출과 다르지 않으나, 추출방식에 있어 더욱 Harsh 한 방식을 사용한다. Tissue를 추출할 경우에는

homogenization이 선행되어야 하며, 그 후에는 세포와 동일하게 진행할 수 있다. 세포 및 homogenized된 tissue를 추출하기 위해서는 ultrasonication, freeze-thaw 방식 등이 가능하다. Ultrasonication의 경우는 간편하며, 추출이 빠르게 진행된다는 장점이 있으나, 강한 진동과로 인하여 샘플에 열이 발생하고, 이로부터 대사의 변화가 일어날 수 있다는 단점이 있다. Freeze-thaw 방식의 경우 용매가 얼었을 때 일어나는 수소결합의 변화에 따라 부피가 커지는 현상을 사용하는 것으로써, 3회의 Freeze-thaw cycle을 액체질소에서 진행하여 대사체를 추출한다. 이의 장점은 추출 자체가 액체질소와 같은 저온에서 진행되기에 대사가 변하지 못한다는 점이며, 단점으로는 반복에 따른 시간효율 저하 등이 있다.

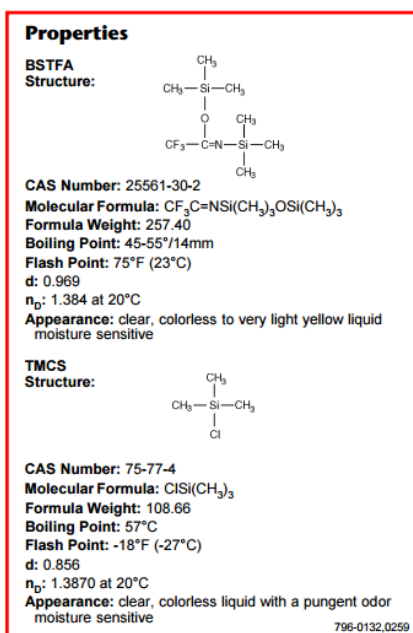
2) 대사체 유도체화

대사체 위와 같이 추출하여 용매를 휘발시킨 뒤 GC-MS에서 검출될 수 있도록 극성 대사체에 휘발성을 부여하는 작업을 진행한다. 유도체화 진행할 때 가장 중요한 점은 시료에 수분이 남아있지 않아야 한다는 것이다. 물은 유도체화 시약의 반응을 가로챌 수 있기 때문이다. 따라서 수분을 확실히 날리기 위해서 동결건조를 사용할 수 있음. 단, 시간이 오래 걸리며, 얼지 않는 유기용매의 경우는 모두 제거한 뒤에나 사용이 가능하기에 상황에 맞게 사용하여야 한다.

대사체 추출물에 1차적으로 Methoxyamine hydrochloride in pyridine을 가하도록 한다. 농도는 주로 20 mg/mL 를 사용하며 시료에 맞게 조정 가능하다. 이 작업은 휘발성 부여가 우선된 것이 아니다. 2차 유도체화인 Trimethylsilylation을 진행할 경우 카르복실기 (COOH)와 히드록시기 (OH) 에 유도체화를 함이 주 목적이거나, 부산물로써 카르보닐기 (CO) 그룹에도 silylation이 진행된다. 그런데 이 카르보닐기의 경우는 Tautomeric form을 가지며, 각 형태에 따라 Trimethylsilylation 이 다양하

게 일어나기 때문에 카르보닐기에 우선적으로 Methoxime을 부착시켜 Trimethylsilylation이 진행되는 것을 blocking하는 역할로써 1차 유도체화 반응을 진행한다.

그 다음으로 휘발성을 부여하는 작업으로써 Silylation을 진행한다. 시약으로는 BSTFA와 MSTFA (Trimethylsilylation), MTBSTFA (t-butyldimethylsilylation) 등이 있으며, 주로 BSTFA, MSTFA를 사용한다. 몇몇 대사체의 경우 유도체화 시약에 따라서 유도체화가 덜되고 더 되는 것은 있으나, 큰 차이는 없기에, 두가지 모두 사용 가능하다. 온도 및 시간 조건은 논문마다 다르나 60도에서는 15분, 45도에서는 30분 이런 식으로 온도가 낮아지면 시간을 길게 하는 방식으로 유도체화를 진행할 수 있다. 또한 각 용액에 Silylation의 촉매로써 TMCS (Trimethylchlorosilane)를 같이 넣어줄 수 있으며, sigma에서는 1% TMCS가 포함된 유도체화 용액을 판매하니 이를 사용하면 된다.



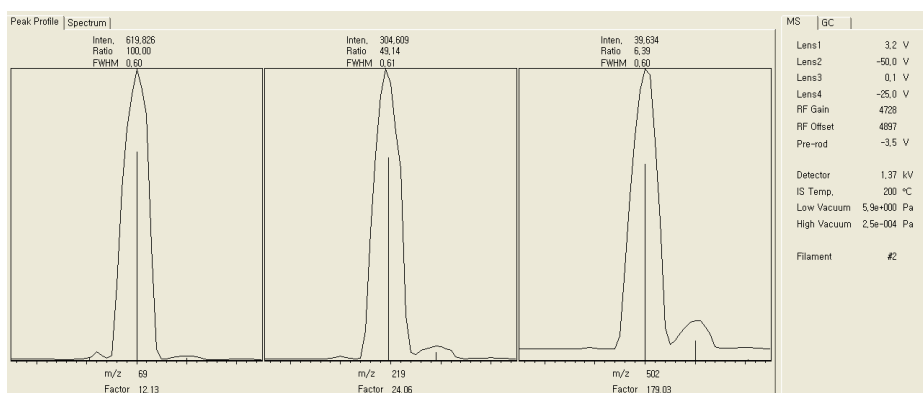
두차례의 유도체화 후, Crimp 형태의 vial에 담아서 가능할 수 있는 대사체의 휘발로 인한 사라짐을 방지하도록 하며, Screw 및 snap 형태의

vial을 사용할 경우라면 data acquisition time을 짧게 하든지 하여 확인 불가한 휘발을 방지하도록 한다. Crimper 및 Decapper는 sigma, agilent 등에서 판매하고 있으며, vial의 size (주로 1.7 mL) 를 고려하여 구매하도록 한다.

3) 기기 사용

기기를 사용하여 데이터를 획득하기 전에 앞서 기기의 상태에 따라서 mass resolution 및 detector voltage를 설정하는 작업이 필요하며 이는 Tuning 에서 진행 가능. 대부분의 기기에서 perfluorotributylamine (PTFBA)를 사용하는 것으로 알고 있으며, 변경도 가능함. Tuning solution은 약 10년간 사용가능하므로, 교체할 일은 거의 없음.

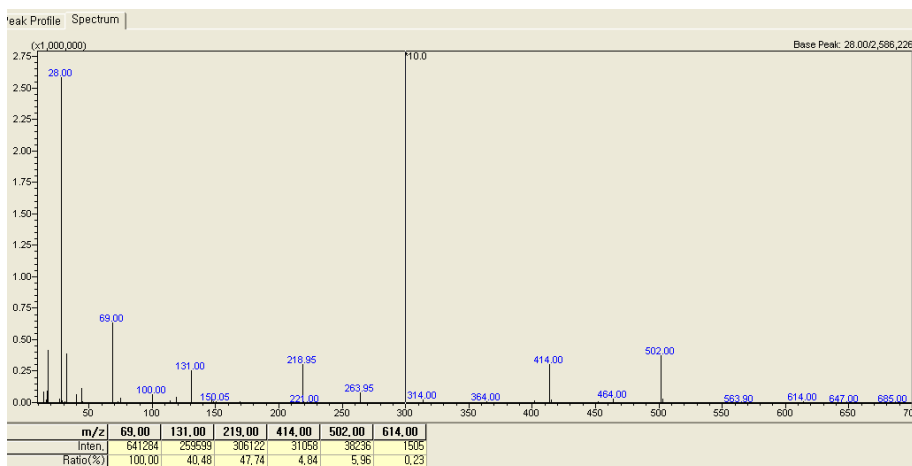
<Tuning result 해석>



@ PTFBA를 사용하면 69, 219, 502 3가지 m/z의 Intensity를 보여주며 그 peak들의 ratio를 보여줌. 이 중에서 High mass (502 m/z)의 ratio가 3을 넘지 않을 경우 Ionization에 문제가 있는 것으로 간주함. Ion source cleaning 을 진행해야 함.

@ 오른쪽 파라미터중 중요한 것은 Detector voltage. PTFBA의 Intensity를 항상 동일하게 맞춰주기 위해서 voltage를 조절함. Ion source가 더럽거나 하여 PTFBA가 적게 유입될 경우 Voltage를 높여 검출 Intensity는

똑같이 유지하는 방식. Detector voltage는 낮게 나타날 경우 기기가 깨끗하고 문제가 없다는 뜻이며, 주로 1.00~ 1.40 kV사이에서 사용 가능. 이 이상으로 높아질 경우 기기에 문제가 있는것으로 간주하고 Ion source cleaning, air leakage 등을 찾아보는 것이 필요함.



@ Tuning spectrum을 보면 28 m/z가 높게 형성되어 있는 것을 확인가능. 이는 공기중 N2 를 뜻하며, 이것이 너무 높을 경우 어느부위에선가 공기가 많이 유입되고 있다는 것을 뜻함. 28 m/z 가 69 m/z에 비해 낮을 경우 적당하며, 조금 높더라도 심각한 문제는 아님. 단, 너무 높을 경우에는 Tuning이 제대로 안되어 detector voltage가 높게 나타남. 이 경우 fitting 부위들을 잘 살펴봐야 하며, 특히 컬럼연결부 중 MS 쪽 부위를 잘 살펴보아야 함.

유도체화가 된 시료가 준비되었을 경우 GC-MS를 이용하여 데이터를 획득하여야 한다. GC-MS에는 다양한 파라미터들이 있으며, 원래는 각 시료의 특성에 따른 파라미터 최적화를 진행하여야 한다. 그러나 대사체 profiling을 하는 경우 흔히 사용되는 파라미터를 기본으로 하여 조금씩 수정하는 방식이 더 적절하다. 기본적인 GC-MS 파라미터 설정은 아래와 같음. 파라미터 중 언급되지 않은 조건의 경우는 조절하지 않아도 되는 경우이다.

<GC-MS method 파라미터 설정>

Sampler | GC | MS

AOC-20i+s

of Rinses with Solvent (Pre-run) : 4

of Rinses with Solvent (Post-run) : 3

of Rinses with Sample : 2

Plunger Speed(Suction) : ☒ High ☐ Middle ☐ Low

Viscosity Comp. Time : 0.2 sec

Plunger Speed(Injection) : ☒ High ☐ Middle ☐ Low

Syringe Insertion Speed : ☒ High ☐ Low

Injection Mode : 0: Normal Sgt...

Advanced...

@ # of Rinses with Solvent (Pre-run): 샘플 Injection 전에 Syringe를 용매로 세척해주는 횟수. 밑의 Advanced 와 연관하여 세척횟수를 설정해야 하며, 용매별로 3회면 충분함. 세가지 용매를 쓰는 경우 꼭 9회로 하여 각 용매 별로 3회 세척을 하도록 함.

@ # of Rinses with Solvent (Post-run): 배치를 이용하여 다수의 샘플을 찍는 경우 다음 Injection에 영향을 끼치지 않도록 시료 Injection 후 세척하는 횟수. 위와 조건 동일

@ # of Rinses with Sample: Syringe에 세척용매가 남은 경우 샘플의 농도가 희석되어 변화할 수 있음. 이에 따른 재현성 저하를 방지하기 위해 Injection 전에 Syringe를 샘플로써 세척하여 세척용매가 잔여하지 않도록 하는 기능. 2회면 충분.

Advanced

Pumping Times : times

Inj. Port Dwell Time : sec

Terminal Air Gap : ☐ Yes ☒ No

Plunger Washing Speed : ☒ High ☐ Middle ☐ Low

Washing Volume : ☒ 6 uL ☐ 8 uL

Syringe Suction Position : mm

Syringe Injection Position : mm

Solvent Selection : ☒ All A,B,C ☐ only A ☐ only B ☐ only C

OK Cancel Help

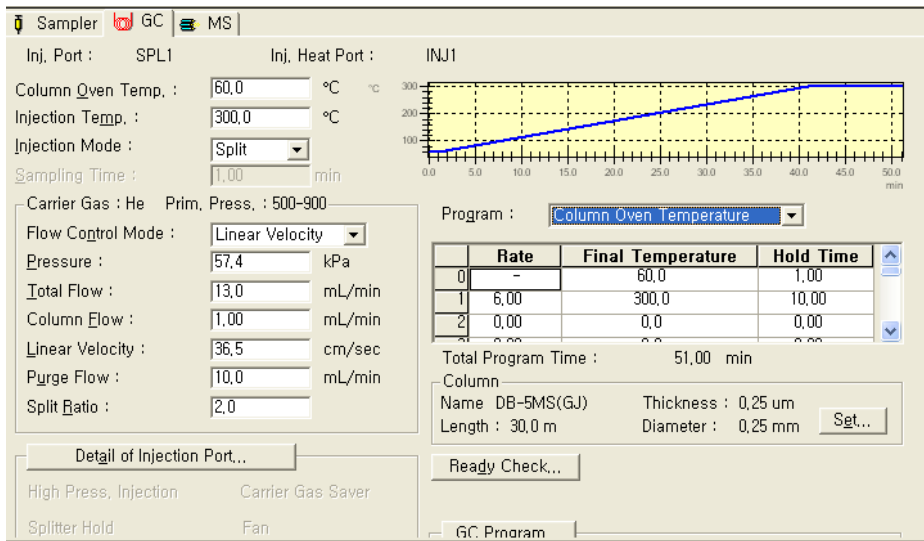
Advance를 누를 경우 나타나는 창임.

@ Pumping times: Sample을 Syringe에 빨아들인 후 잔여한 기포를 제거하는 과정에서 필요한 펌핑 횟수. 3-5회를 사용함.

@ Washing volume: Sample injection volume에 따라서 조절가능하며, 세척용매의 경우 양이 많으면 많을수록 좋음. 8 uL 로 설정하여 사용.

@ Solvent selection: 세척용매 Vial A, B, C 중 선택하여 세척 또는 모든 용매의 세척을 설정함. 용매의 선정이 가장 중요하며, Sample을 모두 제거할 수 있는 세척용매 선정이 매우 중요함. 예를 들어 LLE를 통해 Hexane layer를 취한 뒤 Sample injection을 할 경우 세척용매는 Hexane 또는 그에 준하는 비극성을 지닌 용매를 세척용매로 사용함. 유도체화 된 대사체 추출물의 경우, 극성물질부터 비극성물질까지 다양하게 존재하기에 물+메탄올+메탄올과 섞이는 비극성 용매 (Chloroform, Dichloromethane, Pyridine (Pyridine은 세척으로써는 좋으나 자극적인 냄새로 인하여 사용 잘 안함))의 조합으로 사용하도록 하며, 순서를 물과 비극성용매를 연속하도록 설정할 경우 물과 비극성용매가 섞이지 않아 세척이 제대로 안될 수 있으므로, 메탄올을 꼭 두번째 순서로 놓

도록 함.



@ Column Oven Temp: 오른쪽 Column Oven Temperature program의 시작 온도. 설정은 오른쪽에서 진행.

@ Injection Temp: Injector에서 샘플을 휘발시키는 단계. 대사체 같은 경우 다양한 시료가 존재하기에 250-300도면 적당함.

@ Injection mode: Split, Splitless, Direct 사용 가능. Direct는 실제 데이터 획득에는 거의 사용 안함. 시료 농도에 따라서 Split 또는 Splitless 선택

@ Column flow: Helium 의 유속 결정. 1~2 mL/min 사용

@ Purge flow: Injector부위의 Helium 유입을 추가로 하여 Pressure 조절 등을 원활하게 함. Split ratio가 낮을 경우 Helium Total flow가 10 아래가 되는데 그럴 경우 Purge flow로 Total flow를 10 이상으로 조절해줌.

@ Split ratio: Injection 된 샘플이 실제로 컬럼으로 유입되는 비율 결정. 주로 10 이상으로 사용하나, 대사체의 경우 극미량인 경우가 많아 2로

사용함. Splitless는 사용 권장 안함.

@ Column Oven Temperature Program: Column에의 흡착 정도를 조절하는 기능으로써 온도를 급격히 올리면 샘플이 빨리 나오며, 분리능이 저하되므로 적당히 조절하도록 함. 대사체에서는 주로 5~10의 Rate를 사용하면 분리 및 시간효율 적당함. 주의할 점은 Injection Temperature 온도보다 더 높은 온도조건에 반드시 도달하도록 설정하여 Injection된 샘플이 모두 밖으로 나올 수 있도록 함.

@ Column set: 매우 중요. 실제 장착되어 있는 컬럼과 길이, 두께 등이 설정이 다를 경우 Pressure를 정상조절하지 못하여 기기가 동작 못함. 컬럼 교체 후 반드시 해당 컬럼의 조건으로 변경.

GCMS-QP2010

Ion Source Temp. : 200 °C

Interface Temp. : 300 °C

Solvent Cut Time : 6 min

Micro Scan Width : 0 u

Detector Voltage : ☒ Relative to the Tuning Result ☐ Absolute

Threshold : 100 kV

☐ Use MS Program : Set...

GC Program Time : 51.00 min

Group#1 - Event#1

	Start Time (min)	End Time (min)	Acq. Mode	Event Time(sec)	Scan Speed	Start m/z	End m/z	Ch1 m/z
1	6.00	51.00	Scan	0.20	2500	40.00	500.00	
2	0.00	0.00	Scan	0.00	0	0.00	0.00	

@ Ion Source Temp: 이미 샘플이 이온화되어 있는 상태이기에 온도조건이 많이 중요하지는 않음. 200도 처음 조건 그대로 사용 가능

@ Interface Temperature: 기화된 샘플이 이온화되기 전 통과하는 마지막 통로. 이온화 된 상태가 아니기에 온도가 저하되면 샘플의 액화가 일어나 유입이 안될 수 있음. 따라서 Column oven temperature의 최대 온도 이상의 온도로 설정 해야함.

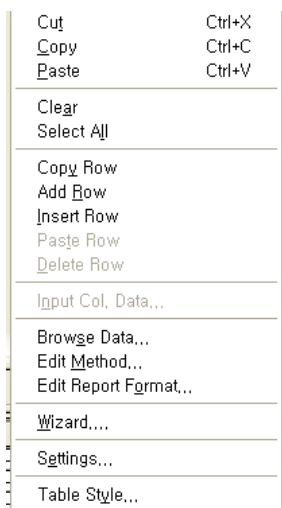
@ Solvent Cut Time: 이온화시키는 기기 소모품인 Filament의 수명을 연장시키는데 중요. 용매도 샘플로써 주입되는데 그 비율이 너무 높으

나 검출의 필요가 없기에 휘발성이 높아 크로마토그램 전반부에서 나타나는 용매 피크는 이온화를 하지 않도록 설정하는 기준. 샘플과 동일한 조건의 용매를 Injection해본 뒤 용매 피크가 나타나는 시간이상으로 설정함.

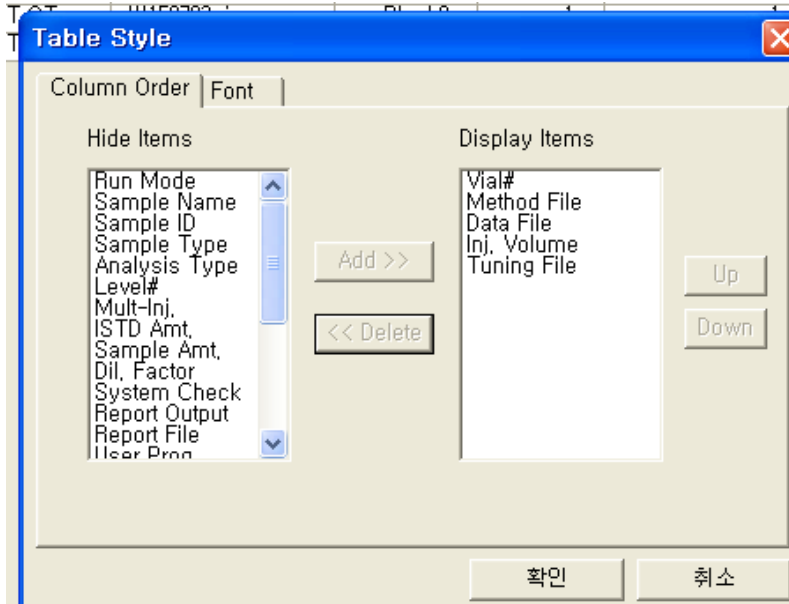
@ Detector voltage: Tuning에 따라서 특정 컴파운드의 Intensity를 동일하게 하기 위한 detector voltage를 설정하는데, 이것에 상대적으로 (relative to the tuning result) 또는 절대값으로 (absolute) 설정하는 것. Tuning시 동일량의 컴파운드를 동일값의 intensity가 나타나도록 조절하는 것이므로 absolute 보다는 relative로 설정하는 것을 권장. 그리고 detector voltage를 너무 낮추면 미량샘플은 검출이 안되며, 너무 높이면 베이스라인이 굉장히 높게 나타나므로, 조절하여 찍어본 뒤 적정값을 사용하도록 함.

<Batch processing>

다수의 샘플을 찍을 때 Batch를 작성하여 자동으로 데이터를 얻도록 함. 아주 간단하며, 몇몇 중요 파라미터만 조절하면 된다. 이는 아래와 같다.

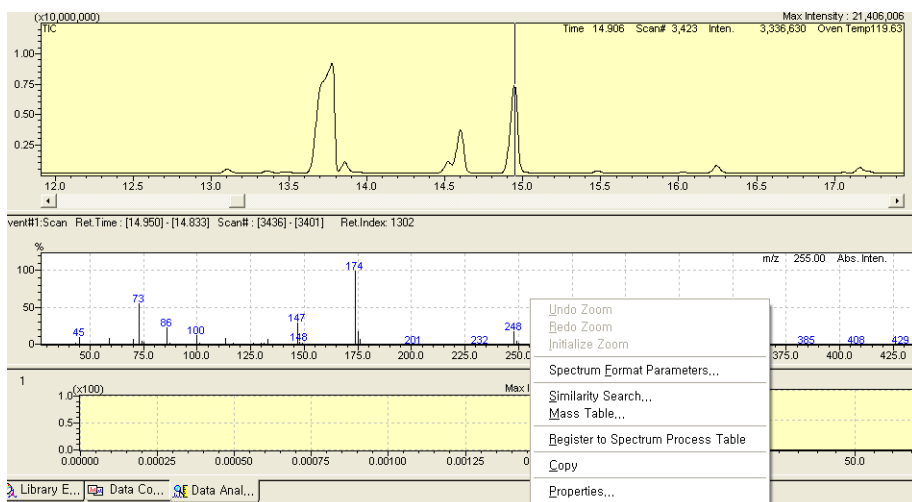


@ 배치 프로세싱화면에서 우클릭 할 경우 Table style이 나옴. 이를 눌러서 필요한 파라미터만 남겨놓을 수 있음. 필요한 파라미터는 아래 그림의 우측부에 있는 것들임.



<데이터 획득 후 대사체 Identification>

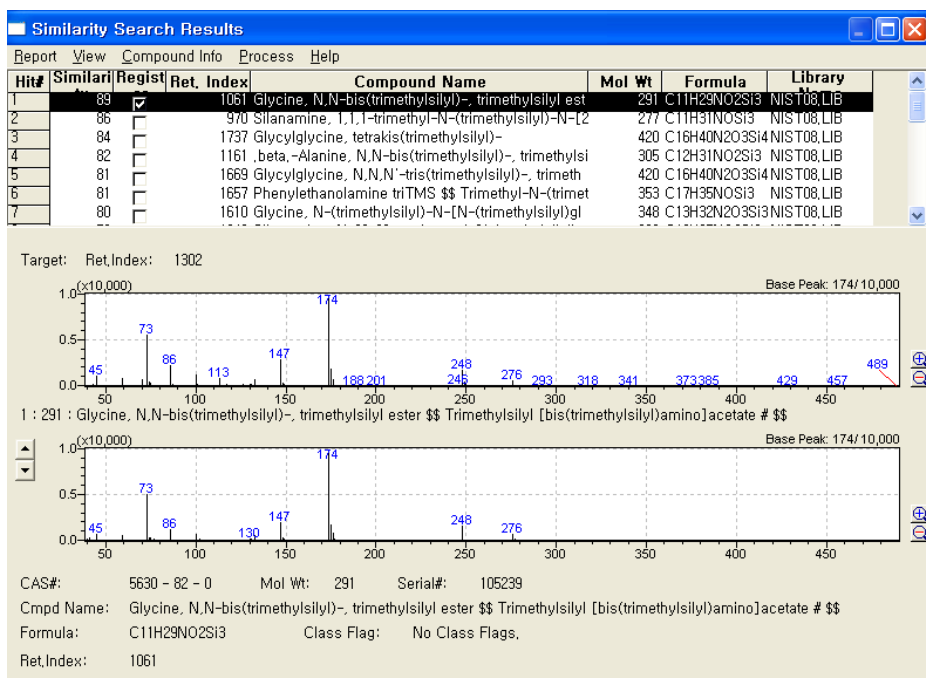
데이터 획득 후 각각의 Peak가 어떤 대사체를 나타내는지를 확인해야 함. 이는 database와 비교를 해야 하며, database는 NIST, WILEY 가 있고, 추가적으로 있는 database도 Shimadzu format에 맞출 경우 사용 가능할 것.



@ 데이터 획득 후 크로마토그램에서 피크의 가장 상단부를 클릭하여 그 부분의 mass spectrum을 아래에 띄우도록 함. 혹시 peak가 너무 작아 baseline에 묻히는 경우 등에는 상단의 버튼 중 Subtract baseline 기능을 이용하도록 함. 아래 그림 중 - 가 붙은 버튼임. 이 버튼을 클릭하여 옆의 베이스라인을 더블클릭할 경우 baseline을 제거한 spectrum을 보여줌



@ 획득한 mass spectrum에서 우클릭을 하여 Similarity search를 클릭하면 설정된 database와 비교를 함. database설정은 상단 Qualitative 메뉴에서 parameters에서 변경 가능.



@ 이때 Similarity index가 80이상이어야 신뢰할 수 있는 것임. 단 80을 넘는다고 모든 Identification이 같지는 않음. 예를 들어 sugar 의 경우는 거의 동일한 spectrum pattern을 서로 가지고 있어서 Identification 할 경우 Mannose와 같은 분자로 모두 동일하게 인식하는 경향이 있음. 또한 80이하인 경우에도 baseline에 의해 묻혀서 확인이 제대로 안된 경우가 있음. 따라서 맹신하지 않고 Mass spectrum을 반드시 확인하도록 함. 또한 Cmpd Name에 나타나는 이름이 우리가 알지 못하는 다른 Synonym일 가능성이 있으므로, 반드시 구조를 확인해본 뒤 유도체화전의 대사체가 어떤것이었는지 확인하도록 함. 예를 들어 Pyruvate의 경우는 전혀 생각치 못한 Cmpd name이 나타남.

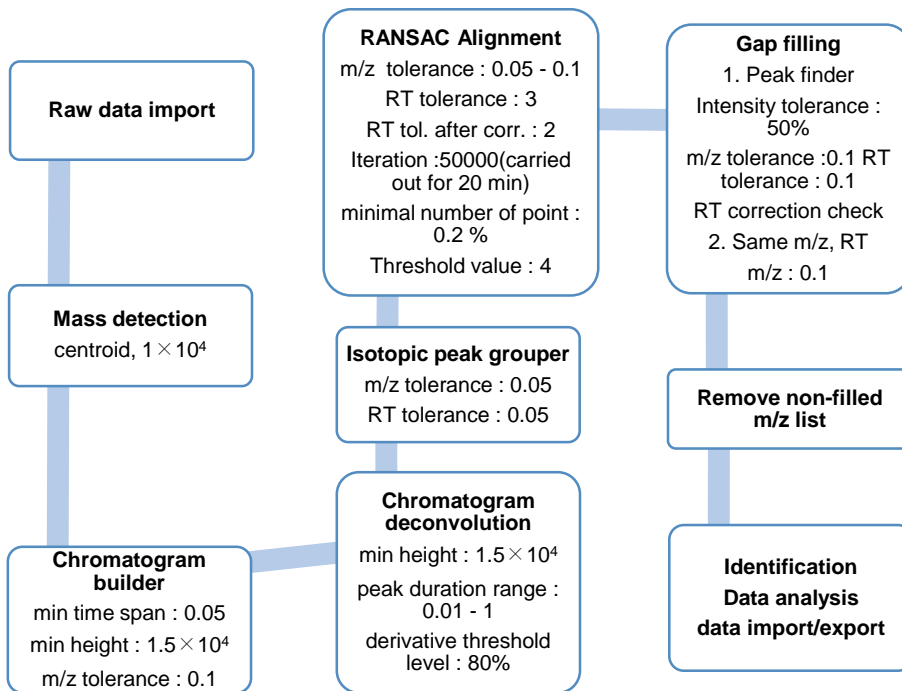
3. MZmine

GC-MS 데이터를 획득한 뒤 통계처리를 하기 위해서는 각 대사체의 Area 값을 기반으로 한 **Bucket table**이 존재해야 함. 이를 획득하기 위해서는 두가지 방법이 있음. 첫번째는 각 샘플에 있는 대사체의 정량값을 모두 직접 얻어 **Table**을 매뉴얼로 만드는 방법임. 이는 시간이 소요되는 면이 있으나, 가장 신뢰할 수 있는 정량값을 얻을 수 있다는 장점이 있음. 두번째는 **Mass alignment software**를 사용하여 **Data alignment**를 진행하는 것임. 이는 파라미터 설정에 어느정도 시간이 소요되어 결국 **Manual**로 정량하는 것과 큰 차이가 나지 않을 수 있으나, 육안으로 보이지 않는 모든 피크들을 획득할 수 있기에 지표대사체 획득에 더 높은 효율성을 가짐.

Mass alignment software에는 여러가지가 있음. 그러나 흔히 사용되는 것으로는 **XCMS**, **MZmine**, **MetAlign** 등이 있음. **XCMS**는 우리가 하는 것이 아니고 데이터를 업로드하면 **SCRIPPS** 센터에서 정렬 및 **Identification**을 직접 하여 건네주는 방식임. 따라서 신뢰도는 높으나, 우리가 스스로 파라미터 등을 조절할 수 없기에 실험자 입장에서 능동적으로 진행하지 못한다는 아쉬움이 있음. **LC-MS**, **GC-MS** 모두 사용 가능. **MZmine**은 가장 **User-friendly**한 software로써 우리가 흔히 접하는 **Graphic User Interface** 방식을 사용함. 따라서 파라미터 설정 시 어려움이 덜함. 그러나, 다양한 파라미터가 존재하고 이를 셋팅하는데는 어느정도 시간이 소요된다. **MetAlign**도 **MZmine**과 유사하나, 파라미터 종류가 **MZmine**에 비해 적은 편이며, 이는 **GC-MS**데이터에만 주로 사용된다. 다만 정렬 후에 선택된 **Peak**들의 크로마토그램 데이터를 획득할 수 있기에, 실제로 어떤 피크가 선택되었는지를 크로마토그램에서 볼 수 있다는 장점이 있음.

그 중 MZmine을 사용하여 해당 논문을 완성하였음. 따라서 MZmine에 존재하는 다양한 파라미터를 어떻게 설정해야 하는지에 대한 간단한 설명을 부록에 넣고자 함.

<MZmine 파라미터 설정>



① Raw data import: 내 데이터를 Import하도록 함. 데이터 포맷은 몇가지를 지원하나, 그중 가장 보편적인 방식인 cdf format을 주로 사용함

② Mass detection: 단순히 Mass spectrum에서 특정 Intensity 이하는 제거하는 방식. 베이스라인에 있는 spectrum중 가장 Intensity가 높은 것을 확인하여 그 이상으로 설정하는 방식과, 내가 보고자 하는 peak 중 가장 낮은 Intensity를 가진 spectrum이 하로 설정하는 방식 둘다 채용 가능.

→ Centroid로 설정하여 획득한 Mass spectrum data를 참고하여 기준선을 결정

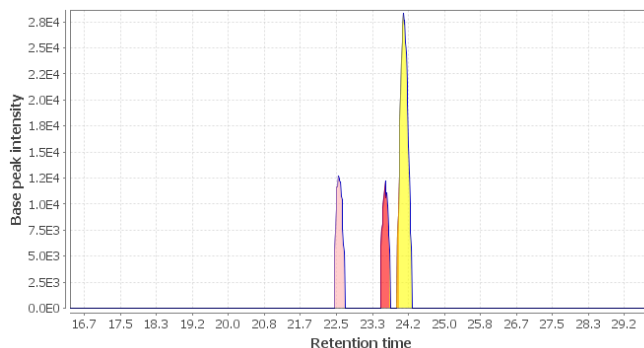
② Chromatography builder

→ Min time span - 특정 m/z 의 스펙트럼의 최소 지속시간 (단위는 분) - True positive와 false positive의 비율 조절

→ m/z tolerance - 같은 spectrum으로 인식하는 최대 m/z 범위 - 획득한 Mass spectrum data 중 특정 m/z 의 변화를 확인 한 뒤 설정

③ Chromatogram deconvolution

→ 서로 다른 compound 중 m/z 가 같은 peaks를 분리하는 기능 - Savitzky-Golay algorithm 사용



→ Minimum peak height - 자신의 peak intensity 에 맞춰서

Minimum peak duration - Minimum peak height의 기준에 맞는 spectrum의 최소 지속시간

→ Derivative threshold level - algorithm 을 거친 뒤 생성된 peak의 Intensity 최소비율 - 80% 면 무난 - savitzky-golay 의 원리 이해 필요 - 창 열렸을 때 밑의 help 를 눌러 참조하면 이해

가 쉬움.

④ Isotopic peaks grouper

- Isotopic peak을 인식하고 monoisotopic만 표시되게 해주는 기능
- m/z tolerance - 같은 spectrum으로 인식하는 최대 m/z 범위
-Chromatogram builder와 같은 parameter 설정
- Retention time tolerance - isotopic peak 들간의 Rt 차이의 최대 범위(단위는 분) - isotopic peak 들간에는 Rt차이 거의없음
- Maximum charge - group 지을 peak의 최대 charge 범위 (metabolomics 에선 charge 2+이상이 흔치 않음)
- Representative isotope - 일반적인 유기화합물(C,H,O,N 구성) 이라면 most intens, lowest m/z 둘다 차이없음, Isotope pattern program 에서 확인한 뒤 선택

⑤ Alignment

- RANSAC Aligner 사용 - algorithm은 help 참조 (간단히 말하자면 Calibration 을 여러 번 반복하는 것과 유사)
- m/z tolerance - 위에서 했던 값과 동일하게 설정
- RT tolerance - 처음 Alignment 시작시 다른 데이터 간 동일 피크로 인식하는 최대 RT 범위
- RT tolerance after correction - RT tolerance 를 기준으로 한 alignment후 동일 피크로 인식되는 최대 RT 범위

→ RANSAC Iteration - 높은 반복횟수를 가질수록 Alignment 결과는 성공적이나 너무 높게하면 아무리 좋은 컴퓨터라도 굉장히 오래걸림, 컴퓨터 사양에 맞게 적당히.

→ Minimal number of point - 각 데이터간 동일 피크의 linear curve를 그리기 위해 사용되는 데이터의 수 (간단히 말하자면, calibration 할때 m/z 의 Error ppm이 높은 기준점은 빼거나, 내가 보고자 하는 m/z에서 벗어난 기준점은 제외하는 경우가 있다. MS의 m/z범위 내 전체 기준점 중 calibration curve를 도출하는 데 사용된 기준점의 비율이 Minimal number of point이다.)

- 0.2% 정도 적당(논문 Comparative evaluation of software for retention time alignment of gaschromatography/time-of-flight mass spectrometry-based metabonomic data 참조)

→ Threshold value - 어떤 기능인지 이해 못함. 4면 적당

⑥ Identification

→ Online database search - KEGG, PubChem, Human Metabolome Database, LipidMaps Database, MassBank, ChemSpider 검색가능. 아직 많이해보지는 않았으며, 이를 위해서는 QTOF와 같은 High resolution mass가 필요할 것으로 보임. 특히 cdf로 변환할 경우 QTOF 데이터도 소수점 밑의 값이 Resolution이 낮아지므로, 처음 획득시 데이터를 그대로 사용하지 않으면 의미가 적음.

⑦ Data analysis, export- Peak area, Peak height 등을 골라낼 수 있음. 이는 실제 엑셀파일을 획득해보면 확인 가능.

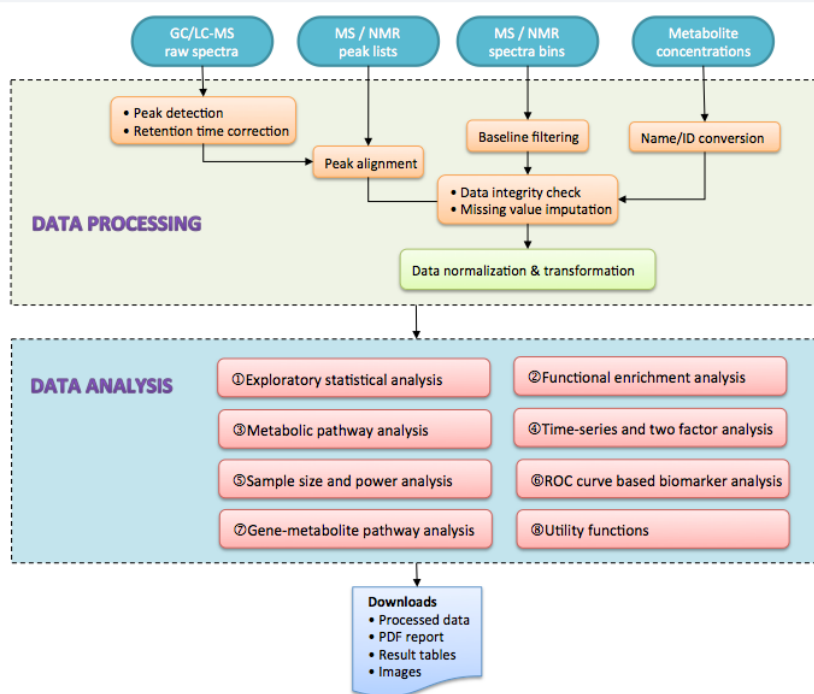
4. Multivariate statistical analysis

한국어로는 다변량통계분석이라고 하며, MS 또는 NMR 등의 기기에서 획득한 데이터는 대체로 대사체 수집 또는 수백개, 더 확장하여 각 대사체에서 나온 수십개의 스펙트럼으로 나누어서 보자면 수천개의 변량 (Variable)이 존재하는 데이터이며, 또한 샘플의 수도 한 그룹당 10~100개까지 존재하여 관측치 (Observation)도 굉장히 많다. 그렇게 될 경우 단변량통계분석인 t-test와 같은 경우는 수천번의 시행을 하여야 하며, 또한 어떤 것이 가장 의미있는지에 대한 판별이 굉장히 어렵다. 따라서 이러한 다변량, 다관측치 데이터를 다룰 수밖에 없는 대사체학의 경우는 다변량통계분석을 기반으로 하여 의미있는 지표대사체를 추출해내는 작업이 필수적이다.

이를 위해서는 다양한 소프트웨어가 사용가능하다. 해당 논문을 위해 사용한 software는 SIMCA-P+ 유료 소프트웨어와 MetaboAnalyst 두가지이다. MZmine에서도 PCA와 같은 간단한 다변량통계분석은 가능하다. SIMCA-P+는 유료라는 가장 큰 단점이 존재하지만, 그래픽을 자신이 원하는대로 수정이 가능하다는 큰 장점이 있다. MetaboAnalyst는 데이터 포맷을 예시로써 제공하며, 홈페이지에서 진행되지만 굉장히 User-friendly하여 사용이 편리하다는 장점이 있다. 그러나 그래픽을 원하는대로 수정하는 것이 불가능하다는 점이 단점으로 지적될 수 있다. 그러나 어느곳에서나 인터넷에 접속만 가능하다면 사용이 가능하다는 점에서 MetaboAnalyst가 가장 유용한 Tool로써 자리매김하고 있는 것은 사실이다. 따라서 MetaboAnalyst를 사용하는 것을 간단히 수록하고자 한다.

Overview:

MetaboAnalyst accepts a variety of metabolomics data types. It consists of two major components for data processing and data analysis. A variety of modules have been implemented to deal with different data types and data analysis purposes. The flowchart below summarizes the overall design flowchart of MetaboAnalyst.



MetaboAnalyst는 Tutorial, Data format등이 모두 제공되기 때문에 적용하는 방법은 굉장히 쉽다. 따라서 어떤 기능인지에 대한 간단한 설명을 넣도록 하였음.

1) Statistical Analysis

➔ Statistical Analysis

This module offers various commonly used statistical and machine learning methods including t-tests, ANOVA, PCA, PLS-DA and Orthogonal PLS-DA. It also provides clustering and visualization tools to create dendrograms and heatmaps as well as to classify based on random forests and SVM.

Statistical Analysis에는 대사체학에서 가장 보편적으로 쓰는 PCA, PLS-DA, ANOVA, T-test 등이 수록되어 있음. 맨 처음에 기본적으로 Data Integrity Check라고 하여 데이터에 대한 Overview를 한번 설명해주며 표에 가끔씩 빠져있을 수 있는 Missing value 를 확인해준다.

<Data integrity check>

크게 중요한 부분은 아님. 혹시나 있을 수 있는 Missing value에 대한 처리를 어떻게 할 지에 대한 선택이 필요함.

Data Integrity Check:

1. Checking the class labels - at least three replicates are required in each class.
2. If the samples are paired, the pair labels must conform to the specified format.
3. The data (except class labels) must not contain non-numeric values.
4. The presence of missing values or features with constant values (i.e. all zeros)

Data processing information:

Checking data content ...passed

Samples are in rows and features in columns

The uploaded file is in comma separated values (.csv) format.

The uploaded data file contains 77 (samples) by 63 (compounds) data matrix.

2 groups were detected in samples.

Samples are not paired.

Only English letters, numbers, underscore, hyphen and forward slash (/) are allowed.

Other special characters or punctuations (if any) will be stripped off.

All data values are numeric.

A total of 0 (0%) missing values were detected.

By default, these values will be replaced by a small value.

Click **Skip** button if you accept the default practice

Or click **Missing value imputation** to use other methods

Missing value estimation

Skip

@ Missing value estimation: Too many missing values will cause difficulties for downstream analysis. There are several different methods for this purpose. The default method replaces all the missing values with a small values (the half of the minimum positive values in the original data) assuming to be the detection limit. Click **next** if you want to use the default method. The assumption of this approach is that most missing values are caused by low abundance metabolites (i.e. below the detection limit).

MetaboAnalyst also offers other methods, such as replace by mean/median, k-nearest neighbour (KNN), probabilistic PCA

(PPCA), Bayesian PCA (BPCA) method, Singular Value Decomposition (SVD) method to impute the missing values ([ref.](#)). Please choose the one that is the most appropriate for your data.

<Data normalization>

Sample normalization

☒ None

☐ Sample-specific normalization (i.e. weight, volume) [Click here to specify](#)

☐ Normalization by sum

☐ Normalization by median

☐ Normalization by a specific reference sample

☐ Normalization by a pooled sample from group

☐ Normalization by reference feature

☐ Quantile normalization

이를 거치고 나면 Data normalization, Data transformation, Data scaling이 존재한다.

@ Data normalization: Normalization은 관측치 (샘플)들간의 차이를 보정해주는 기능으로써 Internal standard 값이 테이블에 존재할 경우 그를 대입하는 것이 가장 좋으나, Internal standard가 없거나 데이터값에 문제가 있을 경우 차선택으로써 by sum, by median 등 등의 기능을 가지고 있음

Data transformation

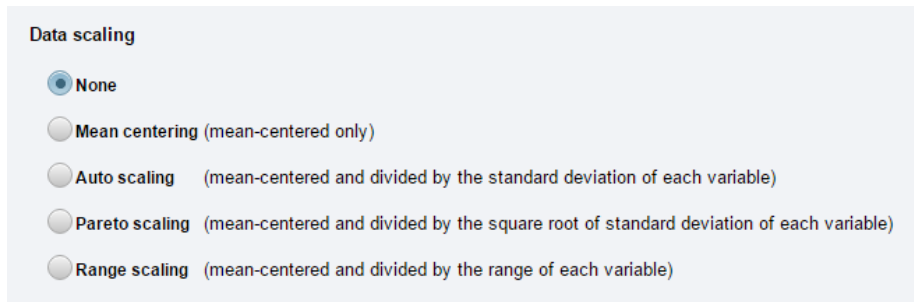
☒ None

☐ Log transformation (generalized logarithm transformation or glog)

☐ Cube root transformation (take cube root of data values)

@ Transformation: 이는 절대값차가 너무 극단적인 변량 사이의 차이를 완화시켜주는 기능으로써 각자 데이터에 맞게 Log 또는

Cube root (세제곱근)을 취하여 수치간 격차 완화를 함. 안사용하여도 무방함.

A screenshot of a 'Data scaling' configuration panel. It features a title 'Data scaling' at the top. Below it are five radio button options. The first option, 'None', is selected with a blue dot. The other four options are 'Mean centering (mean-centered only)', 'Auto scaling (mean-centered and divided by the standard deviation of each variable)', 'Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable)', and 'Range scaling (mean-centered and divided by the range of each variable)'. Each option has a grey radio button icon.

@ Data scaling: 이는 평균값과 편차를 고려하여 중요도의 정도를 고려하여 변량을 보정하여 주는 작업으로써 Intensity가 작아도 흔들림없이 재현성이 높은 대사체의 점수가 더 높아지는 방식의 하나를 생각하면 됨. Scaling의 종류가 다양하게 있으나 이들은 각각 편차의 비중을 어느정도로 두느냐의 차이이며, 데이터에 맞는 scaling을 사용하면 됨

<분석방법의 선택>

Univariate, Chemometrics, Feature Identification, Cluster Analysis, Classification & Feature selection이 있음. 이들중에서 몇가지를 골라서 같이 사용하면 좋으며, 이들에 대한 설명은 생략하도록 함.

Univariate Analysis

[Fold Change Analysis](#) [T-tests](#) [Volcano plot](#)

One-way Analysis of Variance (ANOVA)

[Correlation Analysis](#) [Pattern Searching](#)

Chemometrics Analysis

[Principal Component Analysis \(PCA\)](#)

[Partial Least Squares - Discriminant Analysis \(PLS-DA\)](#)

[Sparse Partial Least Squares - Discriminant Analysis \(sPLS-DA\)](#)

[Orthogonal Partial Least Squares - Discriminant Analysis \(orthoPLS-DA\)](#)

Feature Identification

[Significance Analysis of Microarray \(and Metabolites\) \(SAM\)](#)

[Empirical Bayesian Analysis of Microarray \(and Metabolites\) \(EBAM\)](#)

Cluster Analysis

Hierarchical Clustering: [Dendrogram](#) [Heatmaps](#)

Partitional Clustering: [K-means](#) [Self Organizing Map \(SOM\)](#)

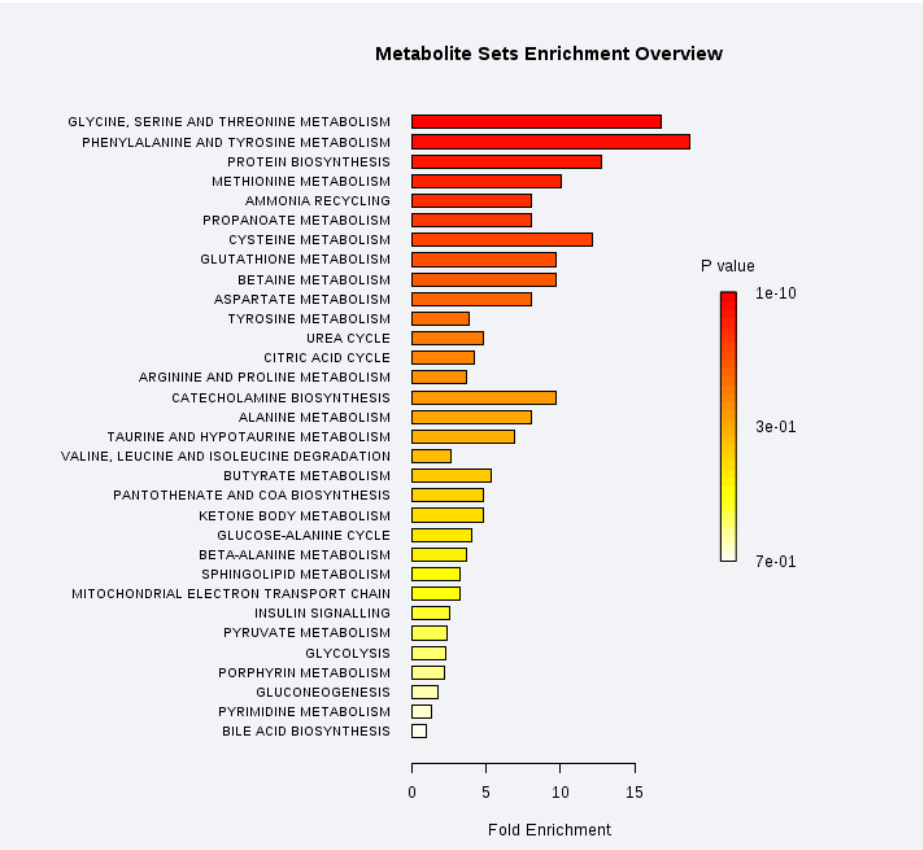
Classification & Feature Selection

[Random Forest](#)

[Support Vector Machine \(SVM\)](#)

2) Enrichment Analysis

이는 얻어진 지표대사체들을 가지고 어떤 대사기전에서 크게 차이가 나타나는지를 찾는 네트워크 분석의 하나임. KEGG pathway 내 존재하는 수많은 대사기전 중 어떤 대사기전이 연관이 있는지를 밝히는 작업을 하며, 그중에서도 대사체의 개수 등이 얼마나 포함되어 있는지에 따라서 p-value를 통하여 연관도 enrichment를 해줌.



	Metabolite Set	Total	Hits	Expect	P value	Holm P	FDR	Details
	GLYCINE, SERINE AND THREONINE METABOLISM	26	9	0.536	1.41E-10	1.13E-8	1.13E-8	View
	PHENYLALANINE AND TYROSINE METABOLISM	13	5	0.268	2.31E-6	1.82E-4	9.23E-5	View
	PROTEIN BIOSYNTHESIS	19	5	0.392	1.94E-5	0.00151	5.16E-4	View
	METHIONINE METABOLISM	24	5	0.495	6.65E-5	0.00512	0.00133	View
	AMMONIA RECYCLING	18	3	0.371	0.00492	0.374	0.0656	View
	PROPANOATE METABOLISM	18	3	0.371	0.00492	0.374	0.0656	View
	CYSTEINE METABOLISM	8	2	0.165	0.0104	0.772	0.119	View
	GLUTATHIONE METABOLISM	10	2	0.206	0.0164	1.0	0.146	View
	BETAINE METABOLISM	10	2	0.206	0.0164	1.0	0.146	View
	ASPARTATE METABOLISM	12	2	0.248	0.0234	1.0	0.187	View
	TYROSINE METABOLISM	38	3	0.784	0.0393	1.0	0.286	View
	UREA CYCLE	20	2	0.413	0.0612	1.0	0.408	View
	CITRIC ACID CYCLE	23	2	0.475	0.0786	1.0	0.484	View
	ARGININE AND PROLINE METABOLISM	26	2	0.536	0.0974	1.0	0.529	View
	CATECHOLAMINE BIOSYNTHESIS	5	1	0.103	0.0992	1.0	0.529	View
	ALANINE METABOLISM	6	1	0.124	0.118	1.0	0.59	View
	TAURINE AND HYPOTAURINE METABOLISM	7	1	0.144	0.136	1.0	0.641	View
	VALINE, LEUCINE AND ISOLEUCINE DEGRADATION	36	2	0.743	0.168	1.0	0.72	View
	BUTYRATE METABOLISM	9	1	0.186	0.172	1.0	0.72	View
	PANTOTHENATE AND COA BIOSYNTHESIS	10	1	0.206	0.189	1.0	0.72	View
	KETONE BODY METABOLISM	10	1	0.206	0.189	1.0	0.72	View
	GLUCOSE-ALANINE CYCLE	12	1	0.248	0.223	1.0	0.81	View
	BETA-ALANINE METABOLISM	13	1	0.268	0.239	1.0	0.831	View
	SPHINGOLIPID METABOLISM	15	1	0.309	0.271	1.0	0.866	View
	MITOCHONDRIAL ELECTRON TRANSPORT CHAIN	15	1	0.309	0.271	1.0	0.866	View
	INSULIN SIGNALLING	19	1	0.392	0.33	1.0	1.0	View
	PYRUVATE METABOLISM	20	1	0.413	0.344	1.0	1.0	View
	GLYCOLYSIS	21	1	0.433	0.358	1.0	1.0	View
	PORPHYRIN METABOLISM	22	1	0.454	0.372	1.0	1.0	View
	GLUCONEOGENESIS	27	1	0.557	0.436	1.0	1.0	View
	PYRIMIDINE METABOLISM	36	1	0.743	0.536	1.0	1.0	View
	BILE ACID BIOSYNTHESIS	49	1	1.01	0.651	1.0	1.0	View

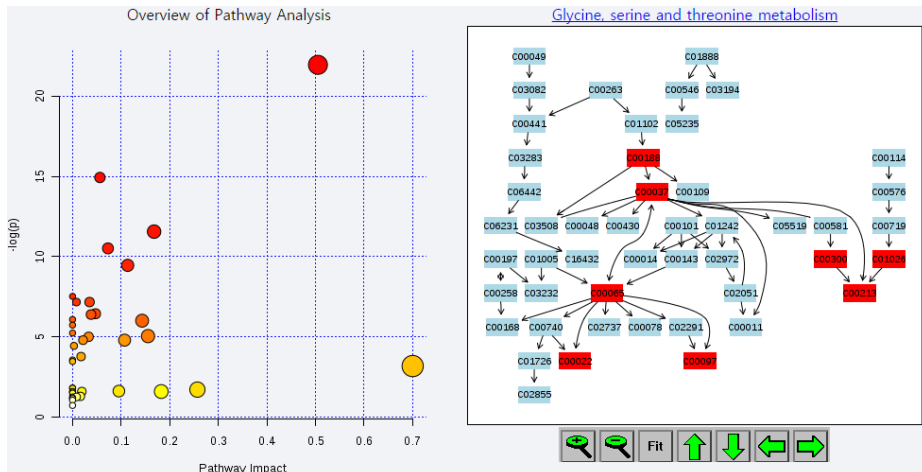
이로부터 실험군에서 두드러지는 대사기전의 변화를 관찰하도록 함.

3) Pathway Analysis

Pathway Analysis

This module supports pathway analysis (integrating enrichment analysis and pathway topology analysis) and visualization for 21 model organisms, including Human, Mouse, Rat, Cow, Chicken, Zebrafish, Arabidopsis thaliana, Rice, Drosophila, Malaria, S. cerevisiae, E.coli. and others, with a total of ~1600 metabolic pathways.

이도 위의 기능과 유사하나, 조금 더 도식화되어 있으며, 또한 각각의 대사체가 해당 Pathway에서 어느 부분에 속해있는지를 확인할 수 있는 기능을 가짐.



4) Time-series/Two-factor Design

➤
Time-series/Two-factor Design

This module supports temporal and two-factor data analysis including data overview, two-way ANOVA, and empirical Bayes time-series analysis for detecting distinctive temporal profiles. It also supports ANOVA-simultaneous component analysis (ASCA) to identify major patterns associated with each experimental factor.

이는 단순히 실험군, 대조군의 성격을 지닌 실험이 아닌 점차적으로 증가하는 (ex, conc 0, 10, 20, 30, Time 1, 3, 6, 10 h) 실험군들에서 유의성있게 증가하는 대사체를 찾아주는 기능을 가짐. 또한 Two-factor design (ex. Control vs disease 와 산사나무 vs 모과

나무 두가지 factor를 동시에 고려) 에서 첫번째 factor가 다른 것
들이 두번째 factor에 의한 차이가 어느정도 차이가 나는지를 비교
해주는 기능을 가짐. 따라서 다른 집단이 특정 약물 또는 처리에 대
한 감응도 등을 비교할 때 사용 가능.

5) 그 외의 기능

Power analysis, Biomarker analysis, Integrated Pathway
Analysis의 경우는 자주 사용되지 않음. Biomarker analysis는 각
대사체에 대한 ROC curve를 얻을 때 종종 사용되며, 특히 임상시
료에 따른 지표대사체 검증시 많이 사용됨. 나머지 기능은 많이 사
용되지 않음.

Abstract

Researchers have long investigated a pathology entailed by parasitism, especially focusing on biological phenomena. But the intracellular response of hosts to the intruders could be a main cause which block the natural interaction between microbe and host. So, we discovered a pathogen-triggered intrinsic activities that induce symptoms under parasitism using correlation network analysis with MS-based metabolomics. As using metabolic pathway enrichment and mapping, we found that the host response to the pathogen changed metabolite levels and thereby accumulated tetrose and pentose sugar alcohols in its host. Next, we performed multiple linear regression model to evaluate the interaction effect between the abnormal accumulation and group variable (control/disease). The statistical values explaining this effect revealed that the accumulation caused a deficient supply of specific sugars, which was found to lead to the stagnation of amino sugar and nucleotide sugar metabolism. The stoppage of this metabolism could hamper pivotal functions for plant host like cell wall synthesis and lesion repair. In conclusion, our findings indicate that correlation network analysis under parasitism could cause stagnation of substrates of the pivotal pathway and, accordingly could trigger pathological symptoms.

Keywords: Metabolomics, Mass spectrometry, Multivariate statistical analysis, Correlation network analysis, Pathogen-triggered metabolism, Sugar alcohol accumulation, Metabolic pathway deficiency,

Student Number: 2011-21751